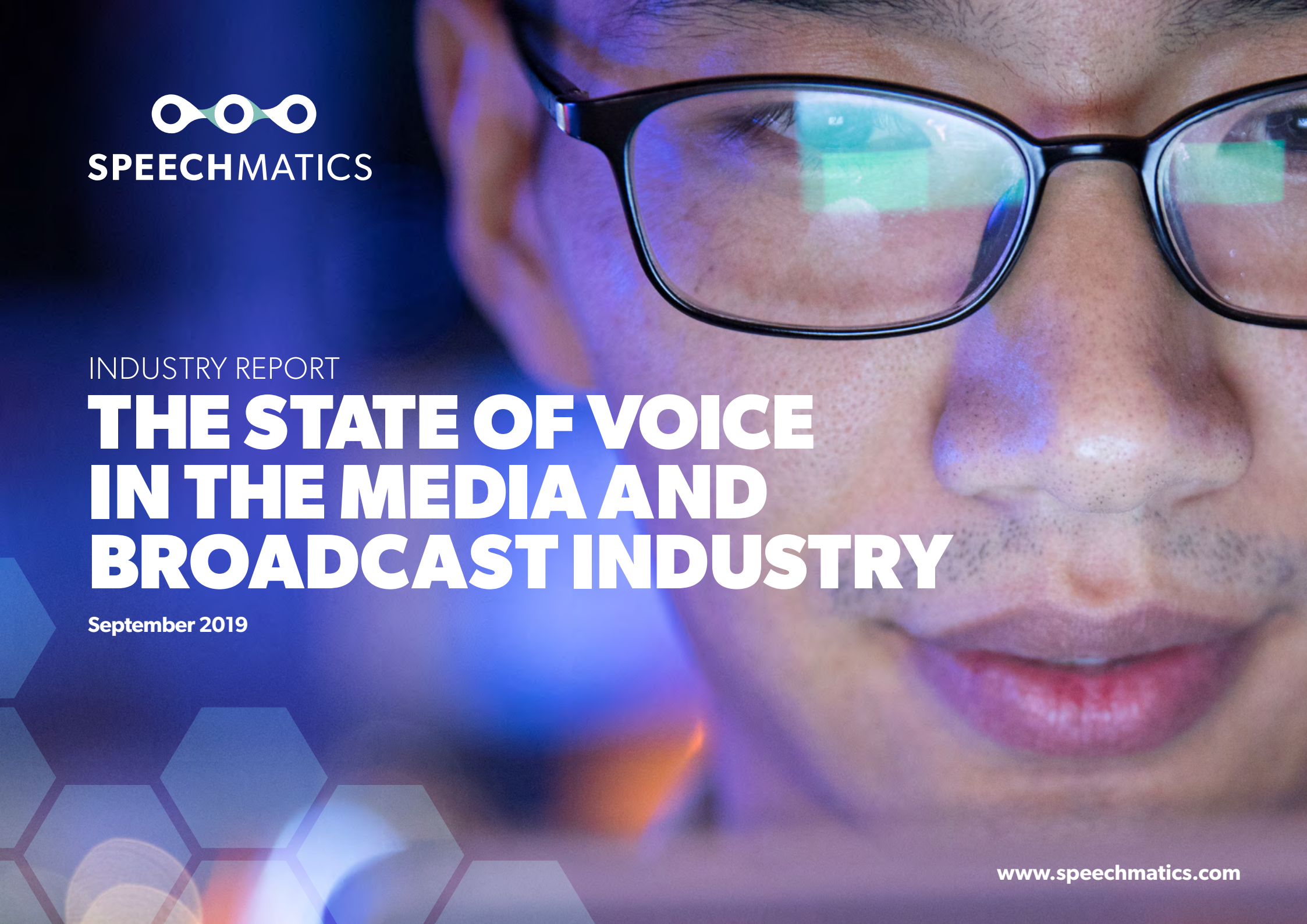


SPEECHMATICS

INDUSTRY REPORT

THE STATE OF VOICE IN THE MEDIA AND BROADCAST INDUSTRY

September 2019

www.speechmatics.com

CONTENTS

EXECUTIVE SUMMARY

Today, the rise in AI and machine learning has opened up opportunities for media companies to expand their capabilities beyond previous expectations.

Voice is now one of the leading technologies for providing added value to the media industry. From media monitoring and media asset management to editing and providing live and closed captioning. The technology helps IT Directors and C-Level IT decision-makers to achieve business objectives.

The results from our research have found that speech technology not only helps to deliver financial rewards but also improves a range of business outcomes, including employee productivity, cost savings, creating a competitive advantage, operational efficiencies and improved customer experience. The research was taken from a range of business sizes from less than 50 up to over 1000, all of which believe there is value in adopting voice technology at the core of their media solutions.

We hope that the findings presented in this report will help to advance the progress of media organisations' voice technology strategies.

FOREWORD

The ongoing growth in hand-held devices means that media is being consumed more and more.

YouTube alone has
1 BILLION+ USERS

GROWTH IN CONSUMER ELECTRONICS

The ongoing growth in hand-held devices means that online media is being consumed more and more, with video becoming the primary channel for consuming information. It's quick, easy and fun to consume and end-users love it. It's no surprise that YouTube alone has over 1 billion users.

Consumer requirements for organisations to provide video content has fuelled the increase in online video and audio content and there are no signs of it slowing down. The demand for media services coupled with the amount of media available has driven the adoption of new technologies and solutions to make media more accessible to the end-user. From optimisation and simplified processes to searchability, discoverability and ensuring conformation to stringent accessibility requirements – media companies are at the core of the step-change in the industry. Making video content more accessible and consumable is vital.

TODAY, THE RISE OF VIDEO

So, what's the situation today? Video is on an upward trajectory across all platforms. Twitter, Facebook, Instagram, YouTube – millions of hours of video are being watched each week across these channels, with companies even expanding their video capabilities to enable users to consume more video content on their platforms. The old adage '*content is king*' has never been more relevant.

By hosting more video content on their platforms, organisations can benefit from maintaining and growing their subscriber base. Revenue growth can be seen through increased subscriptions or generating advertising revenues from attracting large audiences to websites or social media platforms.

Now, media platforms have hit a turning point. Having video content is only a small piece of the puzzle, and having great content is nothing without an audience. The ability for end-users to not only find but to consume and actually **understand** video content has become a challenge. According to GOV.UK, **11 million people** in the UK are deaf or hard of hearing, so how can they understand video without it being accessible through closed captions? It's not only these 11 million people in the UK alone that media companies miss out on either. According to a study by 3Play Media, **"80% of people who use captions aren't deaf or hard of hearing"**. People in public spaces who are situationally disadvantaged often do not want to watch a video with the sound on and would prefer to follow using captions. So how do companies adapt?

THE ROLE OF TECHNOLOGY

Technologies have swiftly evolved to address the challenges of this media content boom. Utilising available technology provides tremendous value for organisations to manage a huge amount of content that they want to make available to their audiences. It also helps to deliver better insights, enhance the customer experience when interacting with the content, and ensure organisations are conforming to regulatory requirements.

The rise of artificial intelligence and machine learning has opened up opportunities for media companies to leverage speech technology. This technology helps to reshape how content is interacted with by the end-user. It opens up accessibility and enables more people to consume information whenever and wherever they wish. Organisations can enhance relationships with customers, employees and other teams within their organisation which is critical to future growth.

The purpose of this report is to explore the state of voice in the media and broadcast industry and provide insights on the true value of voice as determined by professionals within the media and broadcast industry itself. It will address key expectations of speech providers and discuss how organisations can implement new tools to enhance the core of their media solutions.

METHODOLOGY AND DEMOGRAPHICS

For this industry report, Speechmatics collated data points from Owners/Executives/C-Level, Senior Management, Middle Management, Intermediate and Entry Level media and broadcast professionals. These people work in a range of organisations within the media and broadcast industry across the UK, Europe and the United States.

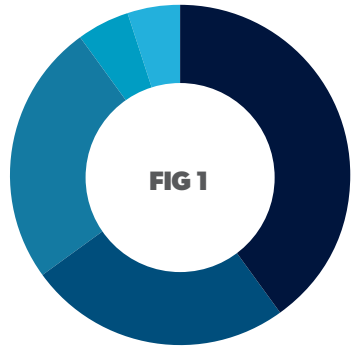


FIG 1. ROLES OF PEOPLE FROM WHICH DATA WAS COLLECTED

- **40%** Owner/Executives/C-Level
- **25%** Senior Management
- **25%** Middle Management
- **5%** Intermediate
- **5%** Other

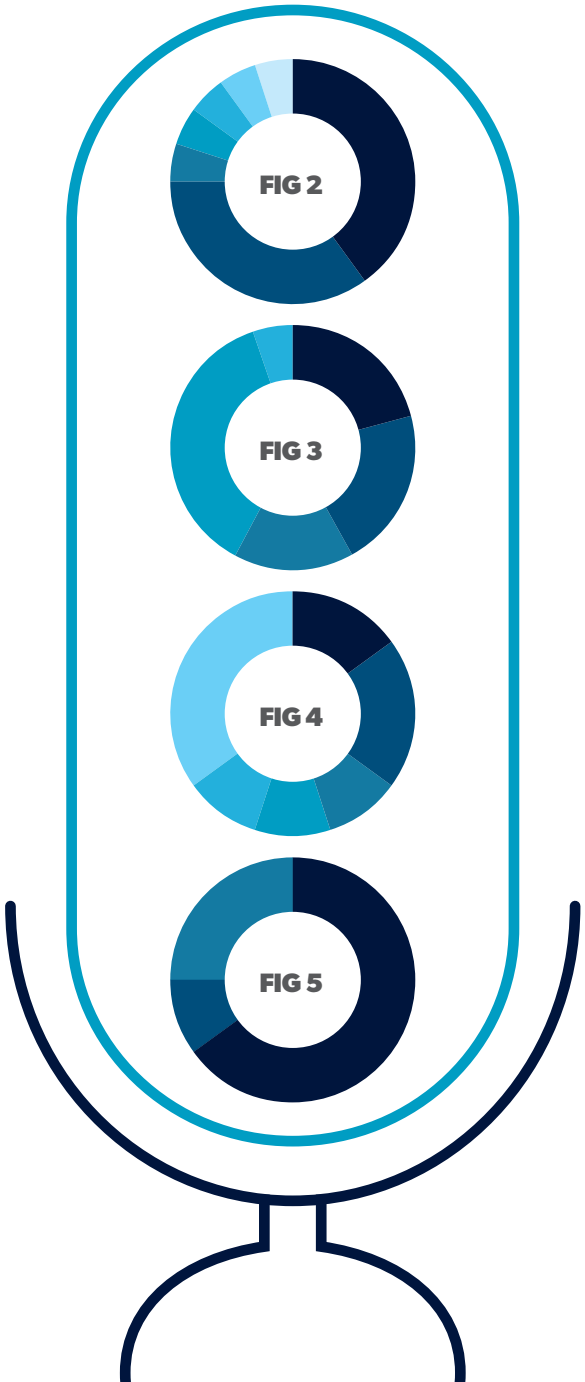


FIG 2. RESPONDENTS DESCRIPTION OF JOB ROLE

- **40%** Management
- **35%** Technical Lead
- **5%** Production
- **5%** Development/Engineer
- **5%** Operations
- **5%** Academia/Student
- **5%** Other

FIG 3. ORGANISATION TYPE IN WHICH RESPONDENT POOL OPERATE

- **21%** Media monitoring
- **21%** Media asset management
- **16%** Editing
- **37%** Subtitling and captioning
- **5%** Other

FIG 4. NUMBER OF EMPLOYEES IN RESPONDENT ORGANISATIONS

- **15%** More than 1,000
- **20%** 501-1,000
- **10%** 251-500
- **10%** 101-250
- **10%** 50-100
- **35%** Less than 50

FIG 5. TYPE OF BUSINESS

- **65%** B2B
- **10%** B2C
- **25%** Hybrid

The respondents described their job roles as Management, Operations, Production, Technical Leads, Developers/Engineers and Academia/ Students and Other. The respondent pool included a wide range of organisations who operate within the media and broadcast industry including, editing, media asset management, media monitoring, subtitling and captioning, and prompting, amongst others.

The collated data encompasses a range of organisations, from large enterprises to smaller start-ups. 15% of organisations surveyed employs over 1,000 people, 20% employs 501-1,000 people, 10% employ 251-500 people, with the remaining 55% employing less than 250 people.

65% of these organisations are business-to-business, 10% are business-to-consumer and 25% a combination of the two.

CONTENT STATISTICS

As the media and broadcast industry continues to grow rapidly, the ability to process the volume of audio and video content increases in parallel. But, how large and rewarding is this growth for media companies? Here are some statistics pulled from WordStream on video marketing.

82%
OF TWITTER USERS WATCH VIDEO CONTENT ON TWITTER

45%
OF PEOPLE WATCH MORE THAN AN HOUR OF FACEBOOK OR YOUTUBE VIDEOS EVERY WEEK

500M+
HOURS OF VIDEOS ARE WATCHED ON YOUTUBE EACH DAY

72
HOURS OF VIDEO ARE UPLOADED TO YOUTUBE EVERY 60 SECONDS

EVERY SECOND, A MILLION MINUTES (17,000 HOURS) OF VIDEO CONTENT WILL CROSS GLOBAL IP NETWORKS BY 2021, ACCORDING TO CISCO (VIA FORBES)

THE 25-34 (MILLENNIAL) AGE GROUP WATCHES THE MOST ONLINE VIDEOS AND MEN SPEND 40% MORE TIME WATCHING VIDEOS ON THE INTERNET THAN WOMEN

YOUTUBE HAS
1BN+
USERS, ALMOST ONE-THIRD OF TOTAL INTERNET USERS

MORE VIDEO CONTENT IS UPLOADED IN 30 DAYS THAN THE MAJOR U.S. TELEVISION NETWORKS HAVE CREATED IN 30 YEARS

500M+
(HALF A BILLION) PEOPLE ARE WATCHING VIDEO ON FACEBOOK EVERY DAY (VIA FORBES)

87%
OF ONLINE MARKETERS USE VIDEO CONTENT

1/3
OF ONLINE ACTIVITY IS SPENT WATCHING VIDEOS

KEY FINDINGS

1

The adoption of voice technology is driven by a need to continually innovate to gain **competitive advantage**.

2

Key **drivers and motivations** for adopting speech technology include:

- Reduced costs
- Productivity improvements
- Support/assistance for human tasks
- Develop better insights
- Operational efficiencies
- Competitive advantage
- Product enhancements

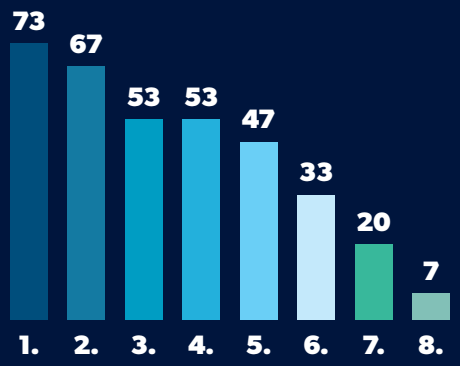
3

Media companies believe that **voice will have a significant impact** on many work-life and home-life activities in 2020, see **Fig 6** below.

4

The adoption of voice technology within the media and broadcast industry offers operational efficiencies, creates competitive advantages, improves customer experience, enables efficient monitoring, auditing and discovery of content.

FIG 6. PREDICTED IMPACT ON MANY WORK-LIFE AND HOME-LIFE ACTIVITIES IN 2020 (%)



1. **73%** Voice search
2. **67%** Enrich AI processes and deliver more personalised interactions
3. **53%** Personal voice assistants
4. **53%** Home automation
5. **47%** Command and control tasks
6. **33%** Bots and automated personal assistants at work
7. **20%** Voice-enabled wearables
8. **7%** Voice commerce

5

Surveyed media companies that have already adopted a voice strategy have seen **increased revenue growth** and **return on their investment**.

6

Media companies find the adoption of voice technology to be **very valuable (60%)** or **valuable (40%)**.

7

Data privacy matters a lot to 80% of the media and broadcast industry.

4

93% of media companies have either considered or are currently **implementing a voice strategy**.

9

Ease of integration is one of the most important considerations when adopting speech technology for the media broadcast industry.

10

73% of media companies use punctuation as a key unit for benchmarking speech technology providers.

11

Language capabilities are considered **very important for 87%** of media and broadcast companies.

MEDIA AND BROADCAST INDUSTRY OVERVIEW

We surveyed media and broadcast professionals from a range of sectors, including:

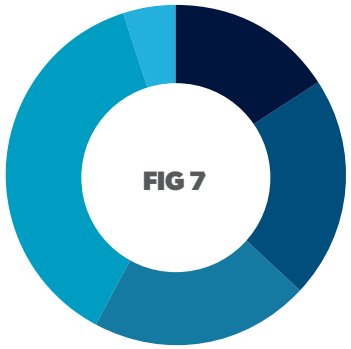
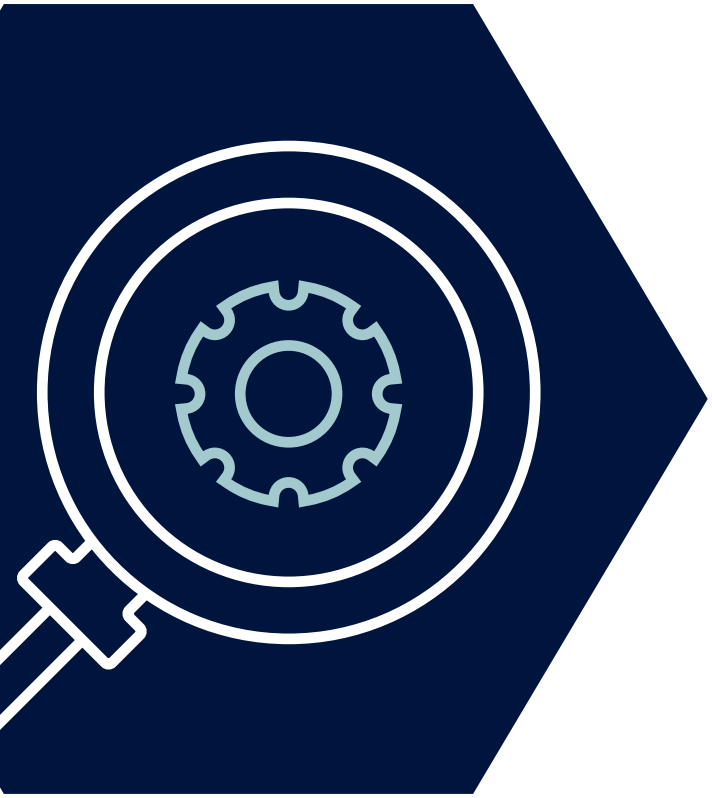


FIG 7. SECTOR FROM WHICH PROFESSIONALS SURVEYED CAME FROM

- **16%** Editing
- **21%** Media asset management
- **21%** Media monitoring
- **37%** Subtitling and captioning
- **5%** Other



MEDIA MONITORING

According to a report from MarketsandMarkets, the global media monitoring software market was approximately $2,260 million in 2018 and is expected to generate around $7,236 million by 2027, at a compound annual growth rate (CAGR) of around 13.9% between 2019 and 2027.

Between 2019 and 2027, the global media monitoring software market has an expected CAGR of around

13.9%

Subsegments of media monitoring include:

- Broadcast monitoring
- Social media monitoring
- Online monitoring
- Print monitoring

Market verticals for media monitoring include:

- IT and Telecommunications
- Retail and Consumer Goods
- Banking
- Financial Services, and Insurance (BFSI)
- Media and Entertainment
- Travel and Hospitality

"At Tedial we have integrated speech-to-text technology within our sports and live events solution SMARTLIVE. We are constantly striving to provide productions team with additional opportunities and editorial value. The technology automatically creates file locators inside Tedial's Evolution MAM allowing operators to search for comments made during a sporting match or live event within seconds."

Jerome Wauthoz, VP Products, Tedial

MEDIA ASSET MANAGEMENT

According to a report by MarketandMarkets the media asset management market is projected to grow from $2.44 billion in 2017 to $5.66 billion by 2022, at an expected compound annual growth rate (CAGR) of 18.3%.

More people are using the Internet each day leading to large amounts of media content being generated. This media content is used and reused by companies across many industry verticals increasing the need to manage this content.

Artificial intelligence and machine learning are transforming how companies manage their digital assets. It is now easier than ever to collate and manage vast quantities of media content, identify specific elements and automatically tag these assets based on certain attributes.

Between 2019 and 2022, the media asset management market has an expected CAGR of

18.3%

While not directly related to the media and broadcast industry, Red Box (a leader in the contact centre industry) commented that **"voice is a critical data set because it holds much more value than any other means of communication"**. Much like in the contact centre market, the media industry is utilising voice data within media content to categorise, index and enhance the discoverability of digital assets based on elements such as keywords, names, people, events, dates, places, genre or other desired categories.

The adoption of voice technology for media asset management companies enables them to significantly improve organisational productivity. It reduces the time taken to search for media clips and considerably cuts costs as a result.

100% of respondents from media asset management companies that have adopted speech technology stated that it was either 'valuable' or 'very valuable' to their organisation. **100% have seen operational efficiencies such as lower costs, reduced time and improved productivity, and 75% said that it gives them the ability to analyse big data sources of audio and video media files**.

"As a business for whom the transcription of language is a key component of our services, speech technology has been transformational for us for well over a decade. In particular, the advances in speech recognition over recent years have allowed us to produce more high-quality captioning for more customers for less cost than ever before – and that's a great result for our customers and their audiences."

Hewson Maxwell, Head of Technology Development, Access Services at Red Bee Media

CAPTIONING AND SUBTITLING

According to a report by MarketWatch, in 2018, the global captioning and subtitling market size was $220 million, and it is expected to reach $370 million by the end of 2025, with a compound annual growth rate of 7.7% during 2019-2025.

Captioning and subtitling comprise encoding, editing, and repurposing of video subtitles and captions for delivery platforms such as web, mobile, and television.

Captioning and subtitling media assets help broadcasting and web media organisations to automate the insertion of captions into a huge amount of audio and video content that is produced daily.

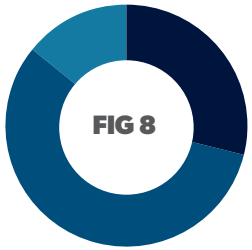
The key driver behind captioning and subtitling is to support accessibility in all forms of communication. As the amount of video content grows, the ability to provide captions for it all becomes more challenging and more costly. On top of this, legislation – in particular, the Federal Communications Commission (FCC) – is moving fast and some tough targets have been put in place for media companies providing captioning for television and online content in the United States of America. Currently, **29% of the market is using human processing** as their solution to captioning, however, the costs are high and require a great deal of human resource to transcribe, align and position captions.

A key reason for the use of human processing is because the media and broadcast market has a very high accuracy demand. Most companies accept a 0–1% word error rate (WER) for most use cases, and at the moment this is only achievable using technology in rare circumstances.

For cases such as broadcast news, 0-1% WER is achievable using voice technology. However, cases that have noisy environments, over talk, multiple speakers, singing or other musical elements may require a range of approaches. **57% of respondents admit to using a hybrid strategy**, leveraging speech technology to deliver the bulk of the work and then human transcribers to provide additional edits.

FIG 8. HOW ARE YOU CURRENTLY DELIVERING SUBTITLES

- **29%** Human transcription
- **57%** Hybrid machine and human transcription
- **14%** Full machine transcription



The captioning and subtitling market has found that voice technology provides huge operational efficiency gains for captioning, reducing the time it takes and the cost to provide the service (83%). The ability for speech solutions to deliver a WER of less than 10% and at a much faster rate than humans for pre-recorded and real-time content, provides significant advantages. Whilst in some instances machine transcription cannot be used in isolation, advances in artificial intelligence and machine learning means that speech technology is on the rise and will be used in conjunction with traditional methods.

100%

of media companies that are utilising speech technology for their subtitling solutions consider voice to be either 'valuable' or 'very valuable' to their organisation

"Speech technology certainly has an increasingly relevant role in caption and subtitle production, particularly where cost and or production time considerations preclude the use of manual processes. This is arguably the situation for high volume, low value or ephemeral content and for live broadcasts, where in essence, the use of speech technology may enable captioning that previous would be uneconomic.

"There are also other applications for speech technology and AI in the quality control, monitoring and archiving workflows, where cost is a significant factor. As a company, Screen Subtitling Systems are actively embracing artificial intelligence-based solutions to support and enable a wider range of workflows and to improve the quality and quantity of subtitle and caption provision in the future."

John Birch, Strategic and Business Development Manager, Screen Systems

EDITING

According to a Report by GlobeNewswire, the video editing software market is projected to grow at a compound annual growth rate (CAGR) of 2.6% to reach $932.7 million by 2025. The growth in video content creation and consumption, as well as increased usage of the Internet and hand-held devices, has fuelled the growth of this market.

Voice technology has been used by media and broadcast companies for some while. Previously, media companies were required to have large teams of editors to edit transcripts and ensure they were accurate. This method was time consuming especially in the cases where a large number of files required checking and editing in parallel.

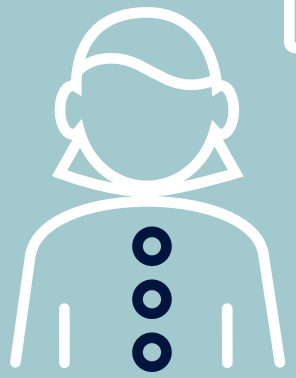
The value of speech technology is no different today to what it was when it was first commercially available. However, the rise of artificial intelligence and machine learning has provided a huge step-change for voice technology, enabling technology to deliver on the promises of WER of less than 5%. The uplift in accuracy from voice technology means that there is less to edit but it doesn't make editing teams redundant. Instead, organisations can make these teams more efficient and their specialist skills can add value to more files than ever before.

"Automatic speech recognition (ASR) technology has revolutionised the way transcripts are created. Creating text from audio was once the exclusive domain of stenographers and typists, but now ASR converts audio to timestamped, punctuated text that can be quickly and easily edited. Additionally, rich features, such as speaker identification, are currently being developed and will add enormous value to ASR algorithm output."

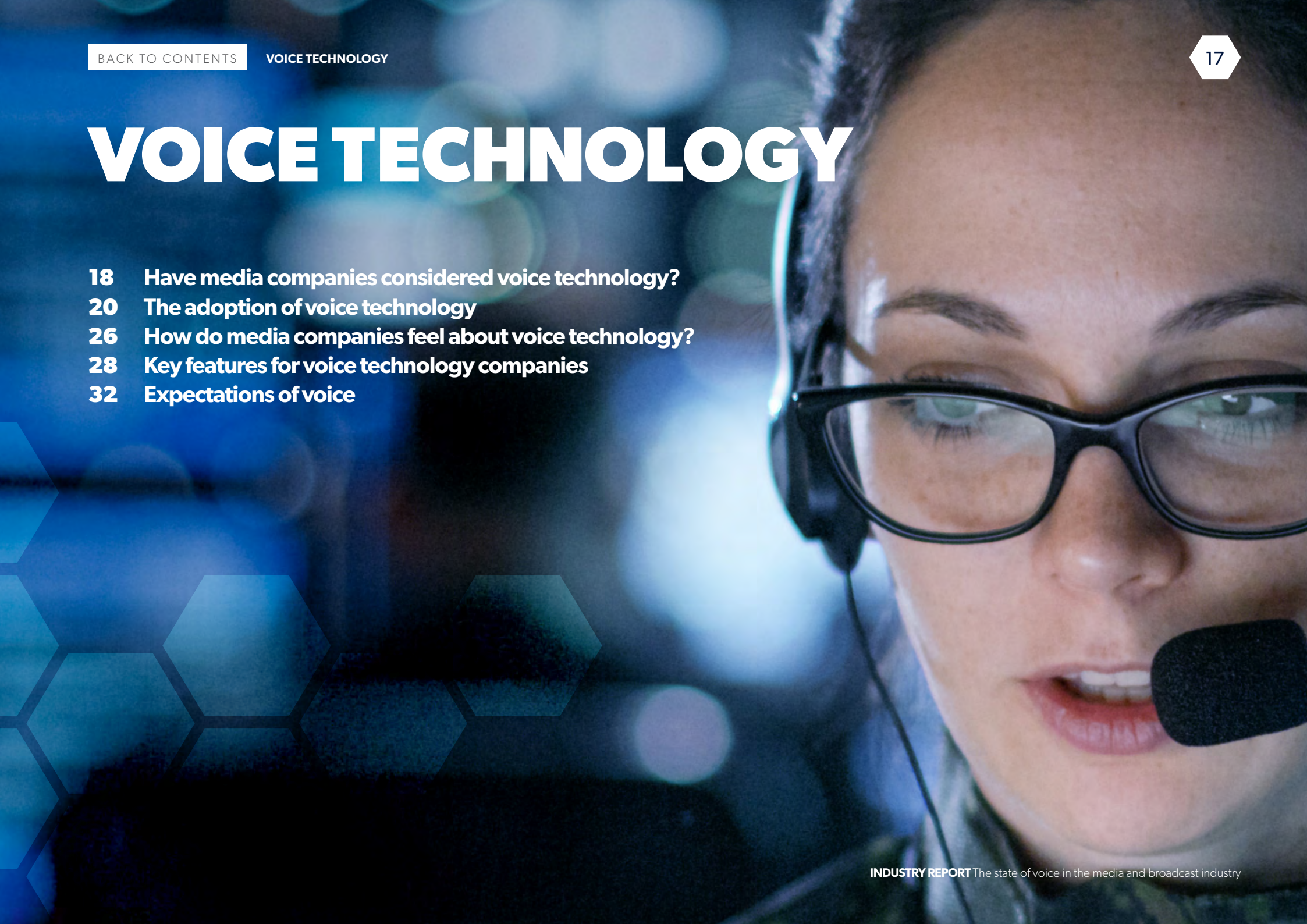
Ellyn Sheffield, **PhD**, CEO and Founder, Verb8tm

Between 2019 and 2025, the video editing software market has an expected CAGR of

2.6%



VOICE TECHNOLOGY



HAVE MEDIA COMPANIES CONSIDERED VOICE TECHNOLOGY?

CONSIDERATION OF VOICE

100%

of media companies surveyed have considered voice technology for their business

IS THERE VALUE IN VOICE TECHNOLOGY?

100%

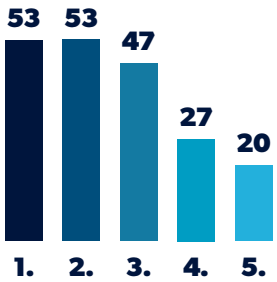
of surveyed media companies that have adopted speech technology find it to be either 'valuable' or 'very valuable' to their organisation

IS VOICE A PRIORITY?

93%

of media companies have either considered or are currently implementing a voice strategy for their business over the next 5 years

FIG 9. CHALLENGES TO ADOPTING VOICE TECHNOLOGY (%)



- **1. 53%** Not yet suitable accuracy level
- **2. 53%** Too complex to deploy speech technology into production
- **3. 47%** Time and resources to integrate the technology into media solution
- **4. 27%** It's too costly
- **5. 20%** Issues with data privacy

CHALLENGES TO ADOPTING VOICE TECHNOLOGY

From our research, 53% of media organisations have already integrated speech technology within their solutions as they believe it generates significant competitive advantage (60%) and better customer experience (33%). A further 20% of companies said that speech technology was a priority for their business in the next 5 years, and another 20% said that they are currently considering the adoption of speech technology.

Integrating any technology comes with challenges and speech is no different. Media companies find that the complexities of deploying speech technology into production is a key challenge (53%), along with the technology not yet being a suitable accuracy level and so a combination of human and machine is required (53%). The complexity of deploying voice technology is a constant challenge that means resources must be allocated to ensure successful integration. It also requires the speech technology provider to ensure that they have processes, procedures, documents, support and training in place to ensure that the deployment process is as easy as possible for their customers.

Just 29% of businesses said that the cost was a key challenge, indicating that the media market is concerned about accuracy and output over price. The concerns around costs are minimal because media companies understand and value that deploying voice comes with operational efficiencies and reduced cost, and so a return on investment is both realistic and achievable.

THE ADOPTION OF VOICE TECHNOLOGY

DOES VOICE TECHNOLOGY INCREASE REVENUE?

40%

of media companies have seen revenue growth since adopting speech technology into their solutions

HAVE MEDIA COMPANIES SEEN A RETURN ON INVESTMENT?

40%

of media companies have already seen a return on their investment since adopting voice technology

WHAT ARE THE BENEFITS OF VOICE TECHNOLOGY?

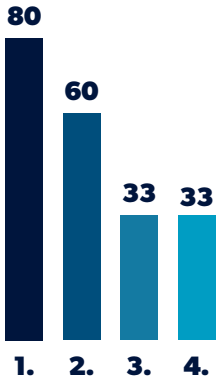


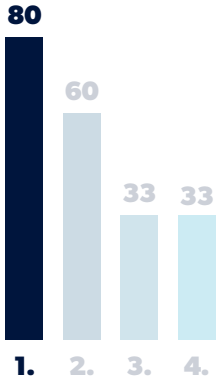
FIG 10. WHAT ARE THE BENEFITS OF VOICE TECHNOLOGY?

1. **80%** Operational efficiencies through reduced turn-around time, lower costs, improved productivity ▶ p. 21
2. **60%** Creating a competitive advantage through new capabilities ▶ p. 22
3. **33%** Improved customer experiences ▶ p. 23
4. **33%** Analysing big data sources of video audio files ▶ p. 24

1

OPERATIONAL EFFICIENCIES THROUGH REDUCED TURN-AROUND TIME, LOWER-COSTS, IMPROVED PRODUCTIVITY

(%)



80% of respondents recognised operational efficiencies as a key benefit of adopting voice technology. The adoption of voice technology enables organisations to process large quantities of content faster than ever before. People often worry that machines will take over in the workplace but actually they are more likely to support employee growth and the quality of work, enriching their working environment. Voice technology is used as a support tool for employees, taking over the manual task of transcribing. Human transcribers can then focus on more skilled editing roles, providing more value to customers.

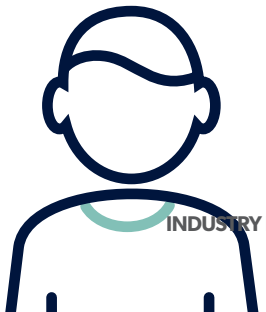
Voice technology not only enriches employees working life, but it also massively reduces costs for businesses. This is important for the media market, with **42% of respondents stating that reducing costs** was a key driver for adopting or considering speech technology.

It was noted that stenographers (people who transcribe speech in shorthand) are becoming harder to come by and are a large expense, and so speech recognition mitigates the challenge of hiring for these types of roles.

"I've been a video editor since 1979. Having time-coded transcripts has forever been an integral tool for my work, but it has been time-consuming and expensive to source.

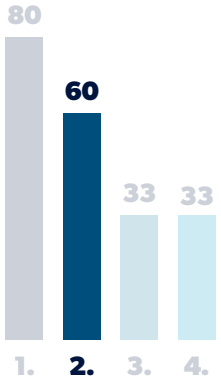
"Having a way to get a fast turnaround of accurate transcripts has been a phenomenal boost to my workflow. It's like magic and I appreciate the access."

Sam Small, Director, Small Wonder Video Services



2

CREATING A COMPETITIVE ADVANTAGE THROUGH NEW CAPABILITIES
(%)



"Our goal is to minimise pain points. Regarding our journalists, they used to spend a lot of time writing down word by word the interviews they had done for radio or TV. These transcripts are the basis for their reports. Now they are no longer dependent on fast handwriting or typing skills but can concentrate on their creative work because the use of accurate working (less than 5% word error rate) speech-to-text software saves them a lot of time.

"More or less the same pain point is solved within our online news team where they remix content from radio and TV and produce web articles; this task is done much faster with the use of speech-to-text. We do use speech-to-text also to generate more metadata of our vast archived content (audio/video). This saves us a lot of time both in terms of archiving content as well as searching for archived content."

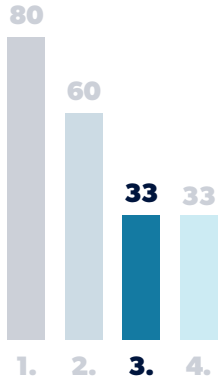
Dr Christian Vogg, Head of Documentation and Archives, Swiss Radio and Television

Media companies recognise the importance of a feature-rich solution that provides them with options to look at expanding their offering into new areas. **60% of companies say that voice technology is beneficial as it creates a clear competitive advantage for their offering**.

Voice technology has already helped media and broadcast companies to **"vastly improve existing products"** and achieve **"business growth and expansion"**. It enables more efficient use of archived media material that was previously inaccessible, expanding capabilities for a range of uses including media asset management and media monitoring.

3

IMPROVED CUSTOMER EXPERIENCE (%)



"Huge productivity gains can be unlocked by using transcription as part of media and research workflows. By making time-aligned words, the key modality for finding quotes and editing unlocks significant opportunities to reduce complexity in challenging and time-consuming tasks, for example cutting quotes together from multiple interviews. The effects are dramatic. Firstly, more people can be part of a workflow. Secondly, transcriptions are suddenly the 'jumping off point' for a series of exciting workflows. And finally, content is now searchable and re-discoverable en masse, whereas before the content was potentially hidden and delivering little value."

Alastair Jardine, Head of Product, Trint

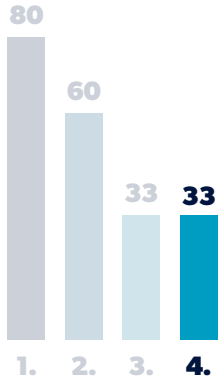
It is a key priority for media companies to provide solutions that enrich their customers' workflows, and voice has been seen to be integral to this offering. **33% of companies stated that improved customer experience was a key benefit** of integrating voice technology within their solutions. It has helped drive better engagement with end-users through the accessibility that speech technology provides. The technology enables users to easily search for and use specific clips from media assets based on keywords, timings, dates etc., to produce better media content.

With time to market a priority for media and broadcasting companies, voice plays a huge role in enabling fast content creation and distribution.

The value of accurate captions and subtitles is already evident. Captions enable accessibility of video content to deaf and hard of hearing audiences. The advancement of speech recognition especially in real-time, means captions are now delivered faster with minimal delay. The addition of advanced features such as improved punctuation characters makes captions even more readable for audiences.

4

ANALYSING BIG DATA SOURCES OF VIDEO AND AUDIO FILES
(%)



Voice technology enables companies to analyse large sources of audio and video files that were previously locked away and very difficult to access. Now, companies can easily access archived material, simply by searching by date, time, keywords etc., to locate specific pieces of content. **20% of respondents mentioned this as a key benefit of voice technology enabling more efficient use of archived material**.

"The process of editing REALITY TV shows is like the proverbial search for a needle in a haystack. Editors must comb through thousands of hours of media to find just a few seconds of interviews worth using.

"The fastest, most efficient way to find the story within this massive amount of information is to work with written transcripts. Unfortunately, traditional human-based transcribing is expensive and often too slow for tight deadlines.

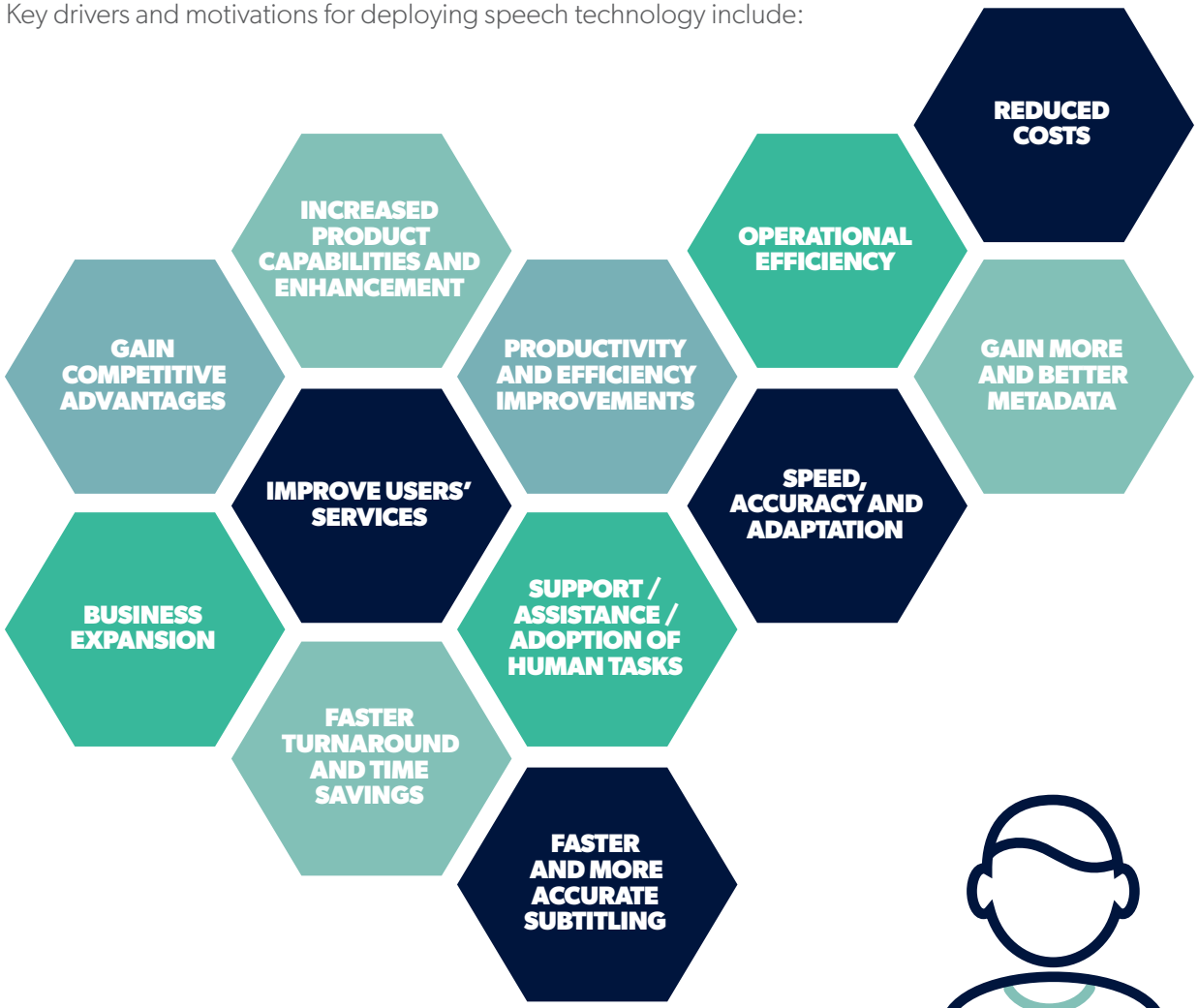
"AI-based transcribing has the promise of being both fast and cheap but is currently less accurate than human-based transcription. As the accuracy rates for AI steadily improve, so does the value of these transcripts for editors.

"Human-based transcription suffers from two fundamental flaws: cost and time. As a result, the amount and type of projects that can use transcripts are limited. Take away cost or time as limiting factors and suddenly a whole new area of use opens up. For anyone editing media with thousands of hours of source interviews, access to low cost, accurate, timely transcriptions is the holy grail. This is what we've been waiting for."

Mark Raudonis, Senior VP Post-Production, Bunim/Murray Productions

WHAT ARE THE DRIVERS AND MOTIVATIONS FOR ADOPTING VOICE TECHNOLOGY?

Key drivers and motivations for deploying speech technology include:



"The rise of VOD (video on demand) and social media created huge demand for language services, and a considerable part of that market can be served by artificial intelligence and machine learning if tuned to perfection.

"Professional producers would never accept garbled, computer-made captions. So, we tuned the solution to a point where it creates subtitles that come close to perfection."

Maarten Verwaest, CEO and Founder, Limecraft

HOW DO MEDIA COMPANIES FEEL ABOUT VOICE TECHNOLOGY?

THE VALUE OF REAL-TIME VS PRE-RECORDED

Our research suggests that **real-time use of speech recognition provides more value (53%)** than pre-recorded (27%), with 20% of respondents identifying both to be equally valuable to their companies. While real-time provides the most value, pre-recorded shouldn't be overlooked as both generate efficiency savings and competitive advantages.

While often subtitles are leveraged for real-time cases, the ability to automatically extract metadata from archived content and enable users to create unique playlists of content delivers huge value with limited effort.

As an example, users can search for specific criteria to generate new content, such as "show me all the times that rain has stopped play in the Wimbledon final since 1998". By searching this term, users would be able to see all the times that the Wimbledon final has stopped because of rain. This can then be used to generate new content.

FIG 11. THE VALUE OF REAL-TIME VS PRE-RECORDED

- **53%** Real-time
- **27%** Pre-recorded
- **20%** Equal

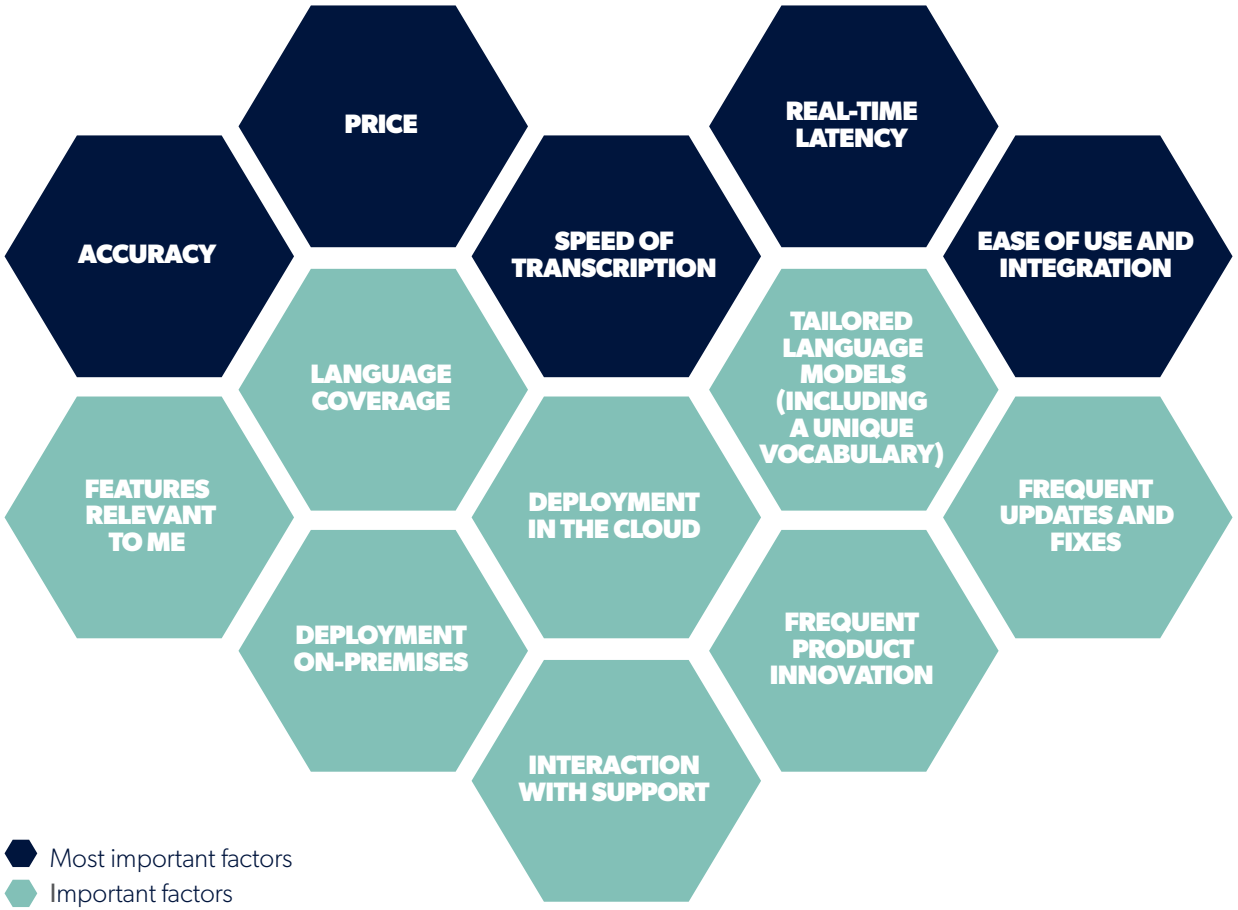
IMPORTANT ASPECTS OF VOICE TECHNOLOGY

According to the media and broadcast industry, the accuracy of speech technology is the most important aspect to consider when choosing a provider.

While accuracy was the most important consideration for voice technology with 93% votes, several other factors were also considered. Unsurprisingly, when comparing speech technology providers, speed of transcription output was also an important consideration as well as price, although price was not seen as a key challenge for adoption. Amongst the list was deployment flexibility (cloud and on-premises), real-time latency, language coverage, ease of integration and relevant features.

With accuracy being featured as the number one consideration, it's important to dive into what this term actually means when it comes to speech technology. As one of the most contested questions for speech-to-text at the moment, we've researched what accuracy **really** means for the media and broadcast industry.

FIG 12. MOST IMPORTANT ASPECTS OF VOICE TECHNOLOGY TO THE MEDIA AND BROADCAST INDUSTRY



KEY FEATURES FOR VOICE TECHNOLOGY COMPANIES

VOICE TECHNOLOGY ACCURACY

With **accuracy occupying either the first or second most important consideration** for speech technology by 100% of respondents, it's clear that it's a critical component. But what does accuracy really mean? From an academic standpoint, word error rate (WER) has historically been how people measure the accuracy of speech technology, but with the increased use of speech technology in the real world, respondents now include more factors in their definition of 'accuracy'.

We asked the media and broadcast industry what accuracy meant to them, the results are shown in Fig 13 to the right.

Whilst WER is important with 87% of people voting it as their definition of accuracy, other factors also contribute towards transcription accuracy and these vary between use cases. The media and broadcast industry indicated that correct use of punctuation and capitalisation, the timing of words, speaker identification, personalised terminology and correct formatting all help to make transcripts more readable and therefore contribute towards the overall accuracy of their output.

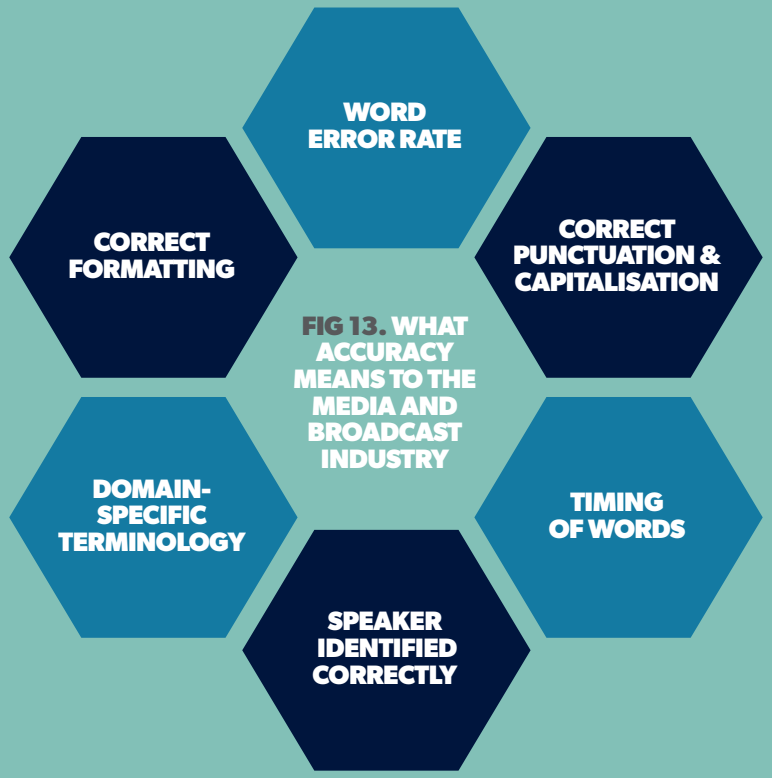


FIG 13. WHAT ACCURACY MEANS TO THE MEDIA AND BROADCAST INDUSTRY

"Pre-broadcast media security is very important for broadcasters/content producers and a few insist on us [Red Bee Media] keeping all media within our network to help guarantee its security. The ability to deploy speech technology solutions locally has been key for that reason."

Hewson Maxwell, Head of Technology Development, Access Services at Red Bee Media

DOES DATA PRIVACY REALLY MATTER?

Not surprisingly, **79% of the media and broadcast industry consider data privacy to matter a lot**, while only **7% say that it doesn't matter**. The other 14% says that it matters a little. Adhering to existing and new industry broadcasting regulations is essential. The main concern for media and broadcast companies is trusting third parties with vast quantities of data that shouldn't be public-facing.

Because the insights held in data can now be accessed easily through speech technology, being able to capture, store and use voice data is now more valuable than ever. Media and broadcast companies understand the importance of insights within voice data and often use on-premises deployment options not only to provide lower latency when transcribing in real-time but also to ensure the security of data.

With popular media like the final series of HBO's Game of Thrones (watched by 17.8 million people, according to CNN Business), the requirements on media companies to keep this kind of media secure until the broadcast date is paramount. The potential impact that a data leak or breach could have on a production company could be millions of dollars. Media companies typically prefer an on-premises solution to ensure that their content and transcripts can remain within a secure environment.

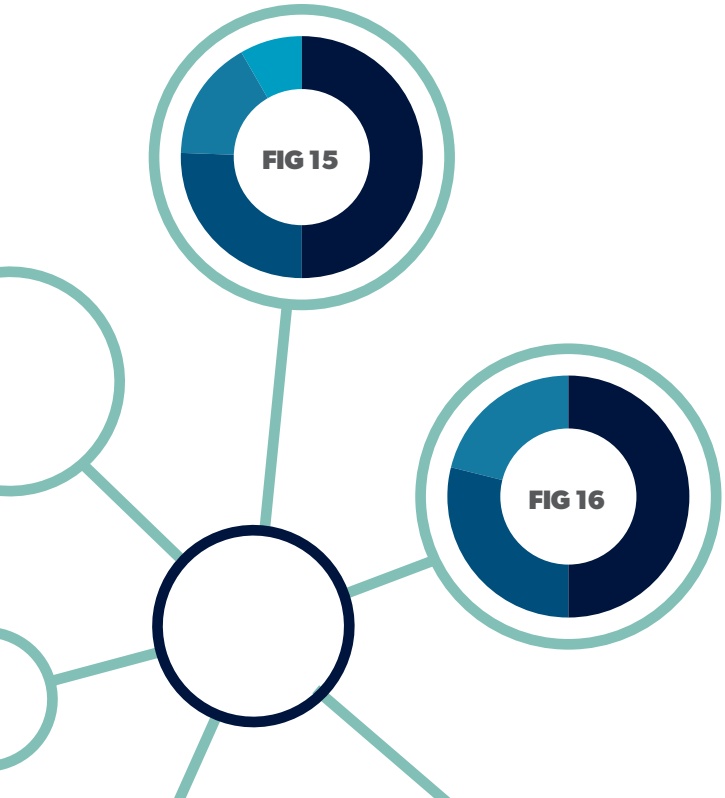
FIG 14. DOES DATA PRIVACY REALLY MATTER?

- **79%** A lot
- **14%** A little
- **7%** It doesn't matter

DEPLOYMENT PREFERENCES

To dive deeper into data security, we asked the media and broadcast industry what most reflects their deployment preference. See Fig 15 below to see what they said.

As shown by the results, 44% of the industry would prefer nothing to leave their own networks. This is not surprising considering how many people value data privacy as very important to their organisations. An on-premises deployment option is therefore extremely appealing for the media and broadcast industry.



EASE OF INTEGRATION

We asked the media and broadcast industry how much help they required to integrate speech technology into their solution, the results are shown in Fig 16 below.

Ease of integration is one of the most important considerations when adopting speech technology for the media and broadcast market. A product that is easy to integrate into an existing solution is essential, especially for cases where the solution consists of many different products layering functionality to deliver the final solution.

FIG 15. DEPLOYMENT PREFERENCES

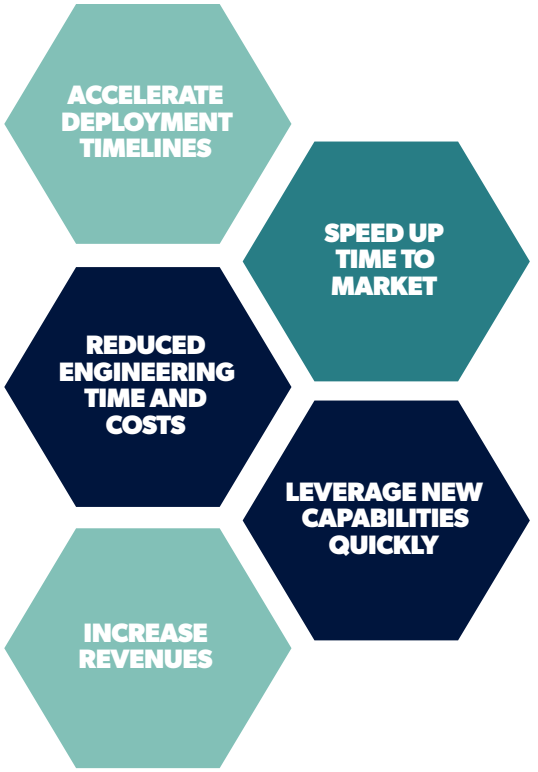
- **43%** Connect to the Internet but nothing to leave the network
- **22%** Use central facility operated by 3rd party but must be single-tenant
- **14%** Can use public shared multi-tenant offerings
- **14%** Privacy not a requirement
- **7%** Must be totally self-contained

FIG 16. EASE OF INTEGRATION

- **50%** Minimal – have own team in-house
- **29%** Some assistance – can use APIs
- **21%** Fair amount

The majority of the media and broadcast industry can integrate products into their solutions using their own teams. However, the ability to provide a simplified integration or even managed services or best practices is valuable for build efficiency and time to market.

FIG 17. KEY BENEFITS OF EASY INTEGRATION



LANGUAGE CAPABILITIES

Languages are considered very important for 87% of media and broadcast companies. Dealing with global companies and operating internationally requires a wide selection of languages, and so this is no surprise.

The most important languages identified are European languages with 80% of respondents calling these out. Specifically, the most important are shown in Fig 19 below.

Other languages identified included, Arabic, Mandarin, Dutch, Japanese, Italian, Portuguese and Polish, amongst others.

As one of the most widely spoken languages in the world, it is no surprise that English was identified as the most popular language. However, popularity comes with the complexities of many accents and dialects. Media companies often select companies that offer multiple language packs for different English dialects. However, this can be time-consuming to run transcripts across all language packs to select the most accurate. An alternative approach is to use global language models that combine accents and dialects as these significantly reduce time costs and provide greater accuracy where multiple speakers are involved.

FIG 18. ARE LANGUAGE CAPABILITIES AN IMPORTANT CONSIDERATION?

87% Yes
13% No

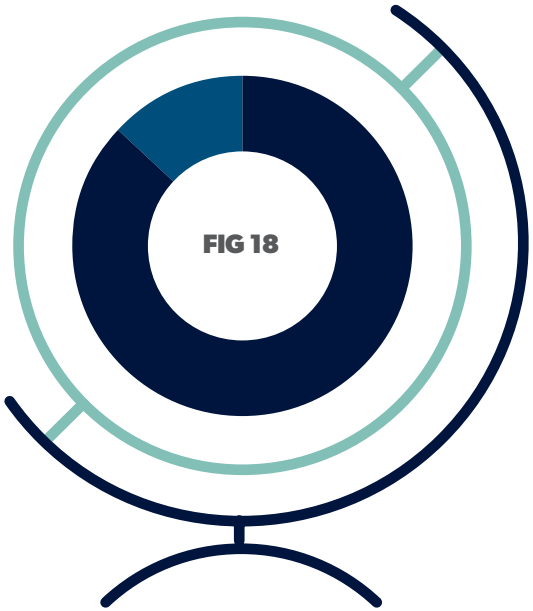
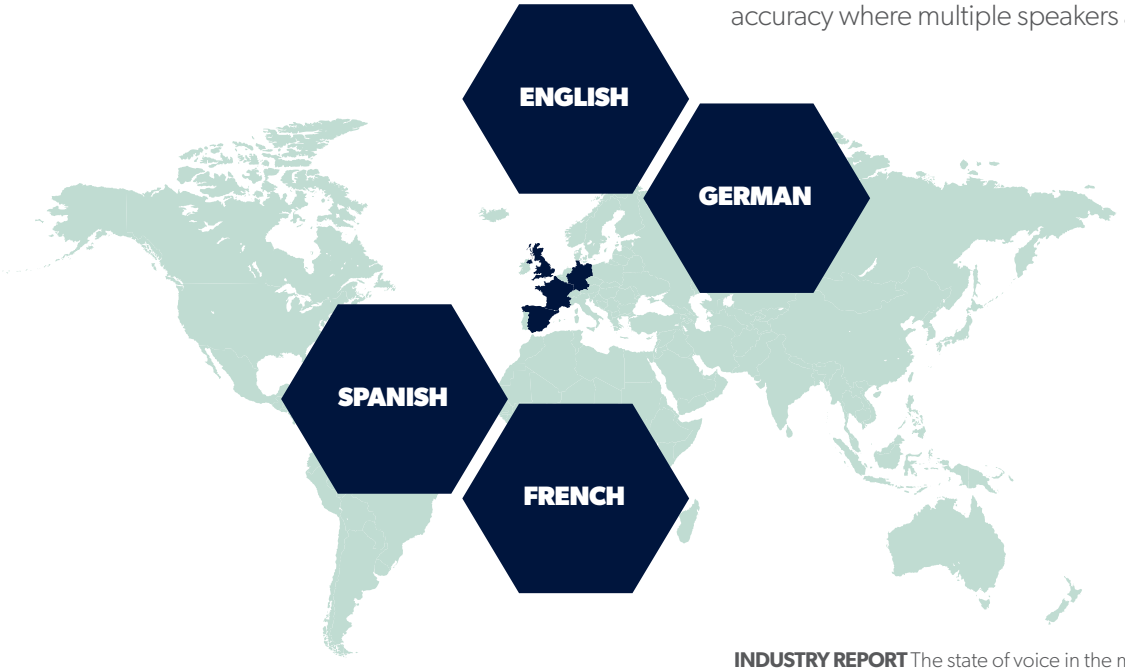


FIG 19. THE MOST IMPORTANT LANGUAGES FOR THE MEDIA AND BRODCAST INDUSTRY

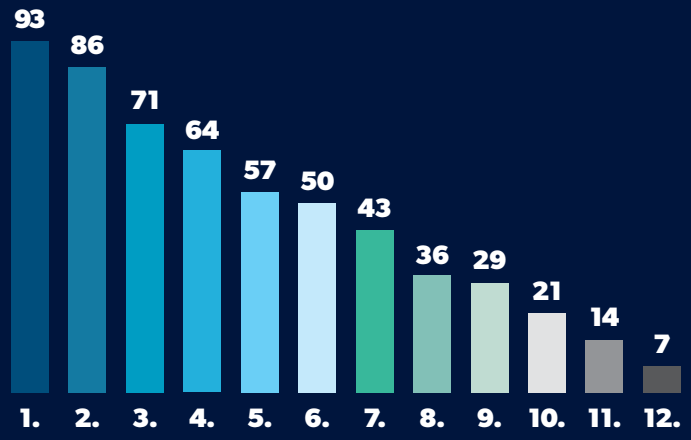


EXPECTATIONS OF VOICE

EXPECTATIONS

The media and broadcast industry expect speech companies to provide a range of features as part of their solution. These features all contribute towards real-world usability and accuracy of transcripts.

FIG 20. WHAT DO MEDIA COMPANIES EXPECT FROM SPEECH TECHNOLOGY PROVIDERS? (%)



1. **93%** Words
2. **86%** Timing, confidence and speaker change
3. **71%** Real-time transcription
4. **64%** Punctuation and speaker identification
5. **57%** Automatic language detection
6. **50%** Alternate words
7. **43%** Non-speech indications, translation and multi-channel audio
8. **36%** Key word detection
9. **29%** Speech editor
10. **21%** Entity extraction and topic extraction
11. **14%** Sentiment
12. **7%** Incorporate bespoke vocabulary

93% of organisations expressed the number one feature that is expected in speech technology is the ability to accurately deliver words from the media that is provided. This isn't a surprise considering that words are the core function of a speech recognition system. Media and broadcasting companies also gave high importance to the timing of words (86%), confidence (86%) and speaker change (86%) in all desired languages. These features enable better usability of transcripts within certain media use cases.

As well as being able to interact with pre-recorded media assets, **71% of organisations expressed that real-time transcription is an expectation**. Media companies are beginning to see the value of voice technology in delivering real-time transcription at the required accuracy and latency.

Speech technology companies now deliver word error rates of less than 10%, and have features dedicated to enhancing readability in many languages. With voice technology, the requirements on humans for editing lessens enabling these resources and skills to be focused on higher-skilled, value-adding tasks increasing revenues and return on investment.

MEASURING VOICE TECHNOLOGY

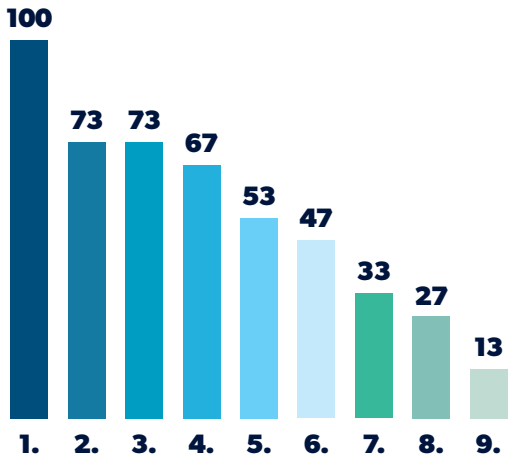
Voice technology delivers significant value for the media and broadcast industry through greater efficiencies, improved product capabilities and delivering a competitive advantage. Companies in the industry regularly benchmark speech technology providers to ensure they are using the most accurate technology for their use case. Failure to select an appropriate provider based on a specific use case could result in significant risk to productivity, speed and quality of the end product and so this process is imperative.

When measuring speech technology providers, **100% of media companies consider correct words as a priority.** Word error rate (WER) is the metric used for determining, roughly speaking, what percentage of the transcript contains errors, and is often used interchangeably with the term accuracy. While WER (and the correct words it measures) is undoubtedly a key component of what respondents care about, other factors help to make transcripts usable for individual use cases. Hence the definition of the word accuracy is extended to encompass more than just that.

73% of media companies also use punctuation as a key unit for measurement. Punctuation enables better readability of transcripts which is especially useful for cases such as captioning and subtitling and accelerating the editing process. 73% also measure turnaround time and speed on transcription when comparing vendors. The results indicate that factors that contribute to the transcript output (WER, punctuation, speaker change) are more important considerations than the price due to the accuracy-driven nature of the industry.

FIG 21. HOW DO MEDIA COMPANIES MEASURE SPEECH TECHNOLOGY PROVIDERS? (%)

1. **100 %** Correct words
2. **73%** Ability to provide punctuation
3. **73%** Turnaround time of transcript
4. **67 %** Speaker change and repeat speaker accuracy
5. **53%** Ease of use
6. **47%** Supported languages
7. **33 %** Secure deployment options
8. **27 %** Speaker identification
9. **13 %** Price



FUTURE OF VOICE TECHNOLOGY

50% of the media and broadcast market agreed that voice is becoming more important and will help to transform how customers interact with their products.

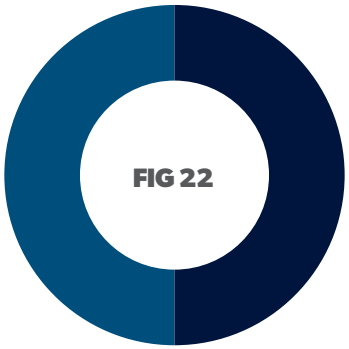


FIG 22. VOICE IS BECOMING MORE IMPORTANT AND WILL TRANSFORM HOW CUSTOMERS FIND AND INTERACT WITH PRODUCTS

- **50%** Yes
- **50%** No

Today, end-users and businesses alike are using more intelligent ways of interfacing with products. The use of keyboards and physical tools are fast being replaced by voice. Intelligent interfaces are combining cutting-edge technologies like artificial intelligence and machine learning with more human-centric engagement methods like conversational voice.

Automated solutions like speech technology can assist with intelligent interactions, with improved accuracy helping to deliver better communication. New features such as punctuation and increased language capabilities uplift the usability of transcripts. With greater accuracy and usability comes greater trust from organisations. Voice technology helps to deliver more insights and relieve manual tasks from the human workforce, enabling them to add value on complex, highly skilled tasks.

Expectations are also shifting for end-users interacting with media assets. With 85% of Facebook video being watched without sound, subtitles are essential for video content displayed on social media. The amount of video content being uploaded is increasing every day so the need for automated transcription has never been more crucial for the accessibility of content.

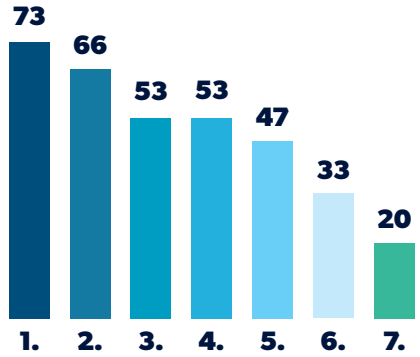
The **top 3** areas that voice will have the biggest impact on according to the media and broadcast market are voice search (73%), enriching AI processes for personalised interactions (66%), home automation (53%) and personal voice assist (53%).

Other areas voice will impact include command and control (47%), bots and automated personal assistants at work (33%) and voice-enabled wearables (21%).

Transcribing media assets delivers lots of value to media and broadcasting companies. Converting media files into text delivers flexibility to support a wide range of use cases and improve customer interactions and experiences.

Voice search not only requires an accurate transcription but also the capability to tailor voice models to include uncommon words such as names, places, acronyms and custom vocabulary. With more content creation comes increased importance on efficient and automatic indexing. This is vital to ensure that users can find and access the content that they want. What better way to generate this metadata than the content itself?

FIG 23. AREAS VOICE WILL HAVE THE LARGEST IMPACT ON (%)



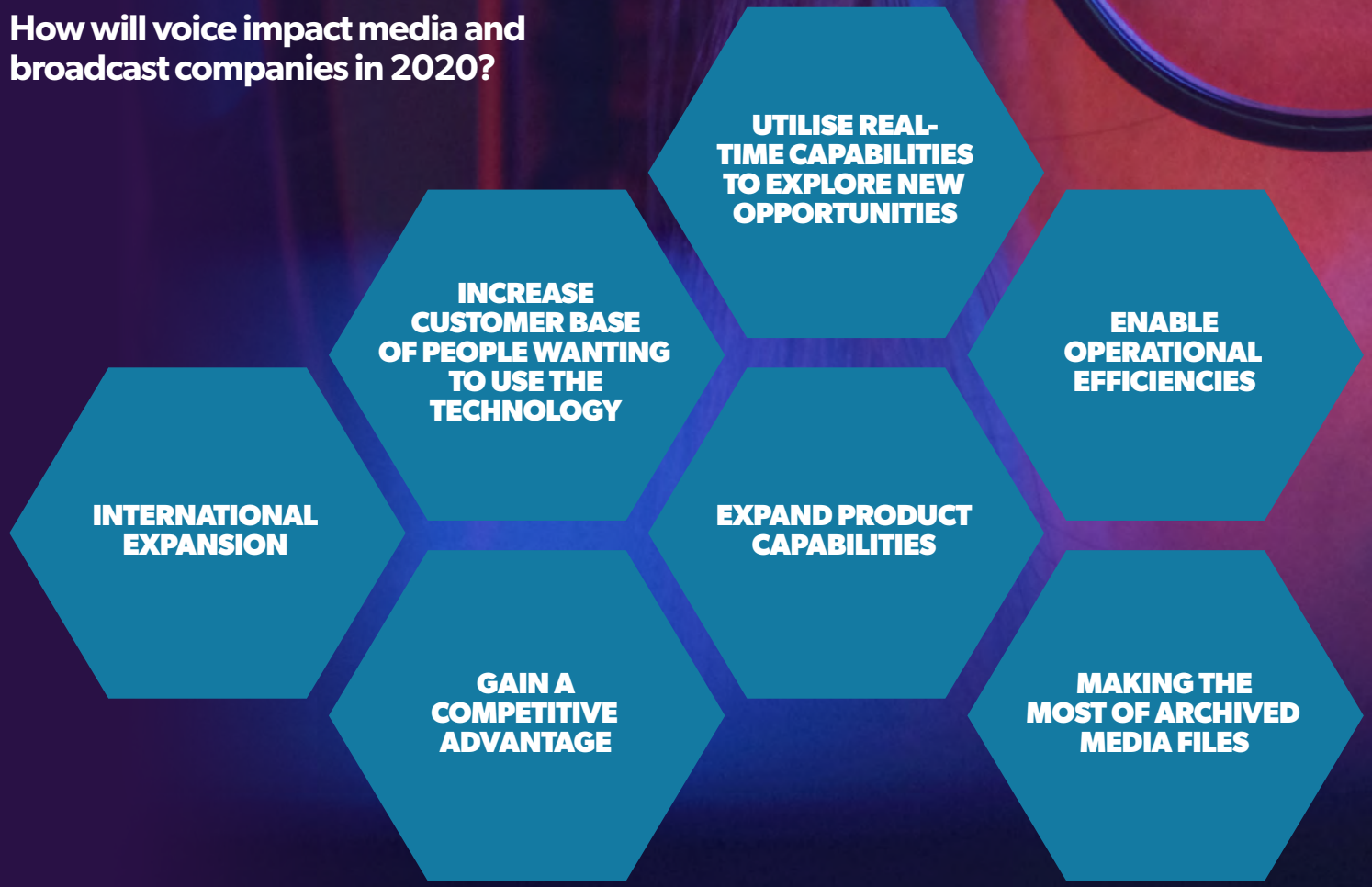
1. 73% Voice search
2. 66% Enrich AI processes for personalised interactions
3. 53% Home automation
4. 53% Personal voice assist
5. 47% Command and control
6. 33% Bots and automated personal assistants at work
7. 20% Voice-enabled wearables

WHAT DOES THE 2020 MEDIA AND BROADCAST INDUSTRY LANDSCAPE LOOK LIKE?

Predictions for the use of voice technology in 2020 from the media and broadcast industry.

WHAT DOES THE 2020 MEDIA AND BROADCAST INDUSTRY LANDSCAPE LOOK LIKE?

How will voice impact media and broadcast companies in 2020?



SUMMARY

The adoption of consumer electronics with the capability to capture high-quality video content has helped fuel the growth of the media and broadcast industry.

Production companies are continually battling for audience growth and so personalisation and accessibility is on the rise. The way that content is consumed is changing from linear to on-demand and because of this, media and broadcasting companies require more sophisticated methods of extracting metadata within content. This has seen a rise in demand for better indexing and searchability of digital assets.

The adoption of speech technology for the media and broadcast market is becoming essential for companies to deliver on their promises. The rise of artificial intelligence and machine learning has been instrumental in the increased adoption of speech technologies. Accuracy continues to be the key metric when it comes to evaluating speech providers due to the time and cost-saving this brings.

Continual reductions in word error rate is, therefore, opening up more applications for the technology. As speech technology improves, media companies are now placing more trust in AI and speech technologies to automate their processes, reduce manual effort and generate a return on investment of deploying speech technology within the core of their media solutions.

The benefits of speech technology within the media and broadcast industry are undeniable. 100% of media professionals surveyed said that it was either 'valuable' or 'very valuable' to their organisations. From accelerated transcription and reduced dependencies on human transcribers to extracting valuable data from within video and audio content, the media and broadcast industry is seeing efficiency gains and opportunities for product enhancements and business expansion. Digital assets are managed and indexed effortlessly, meaning they are easier to find by the end-user, ultimately improving the customer experience.

Speech technology is continuing to evolve at an increasing rate and closing the gap on the 0-1% word error rates required by the media and broadcast industry. While voice technology has already proven to provide value through automatically transcribing media assets and removing the heavy lifting from human transcribers, it is not yet perfect. People will continue to be an integral part of the media workflow. Advances in voice technology will mean that organisations can automate large quantities of manual transcription work. Organisations can then better utilise their workforce by augmenting their responsibilities to add value outside of the low-skill tasks and add value to activities currently out of reach by automated processes.

With the rise in artificial intelligence and machine learning, there hasn't been a better time for media companies to adopt speech technology to enhance the core of their media solutions.