SPEECHMATICS

INDUSTRY REPORT

TRENDS AND PREDICTIONS FOR VOICE TECHNOLOGY IN 2021

JANUARY 2021

WWW.SPEECHMATICS.COM

CONTENTS

EXECUTIVE SUMMARY

2020 was a year like no other and few could (or would) have predicted the COVID-19 pandemic that still impacts the globe today. The results of the Speechmatics 2021 trends and predictions report reflects the impact of COVID-19 as organizations looked to leverage voice technology.

The introduction of voice technology into business workflows could fortify product offerings and services to meet the challenges created as a result of the virus. In some cases, the actions of organizations and industries were dictated by the reactions to initiatives like social distancing and working from home to curb the spread while trying to maintain effective working practices and business stability.

As a result, the application of technologies such as voice are being considered even more important as countries, industries, organizations and individuals continue to operate in a world that is now defined by COVID-19 and look to plan for a post-pandemic world.

This report will explore the history, development, expectations and trends when it comes to solutions like voice technology. It will also look at the impact that COVID-19 has had on the automatic speech recognition (ASR) market specifically, and other industries in which voice technology is utilized.

It contains key insights from industry experts, product specialists and machine learning engineers at the bleeding edge of these technologies. They reveal their opinions and expectations of voice technology, the markets it will influence, the benefits it will have and the capabilities it will enable.

The report will also look into how machine learning and AI specifically influence the ASR market. It will address the value and benefits that voice technology can deliver as it continues to evolve and mature, as well as the drivers and motivations behind its adoption. The report will cover the challenges of adoption and key elements that are required to see continued growth.

FOREWORD

Today's world and workplaces are almost unrecognizable from how they looked and operated in 2019 and early 2020. When the pandemic hit, while the impact was massive from a digital point of view, many businesses were in a pretty good shape. Web conferencing technology was already a core business collaboration tool – as well as other digital technologies – which meant businesses could quickly plug the gap that was left with the removal of face-to-face interactions. Voice technology has helped to maintain productivity and operational capabilities as the world changed around them. Core voice technology – like the ability to accurately transcribe voice – not only makes conversations visual but accessible and unlocks key data for businesses.

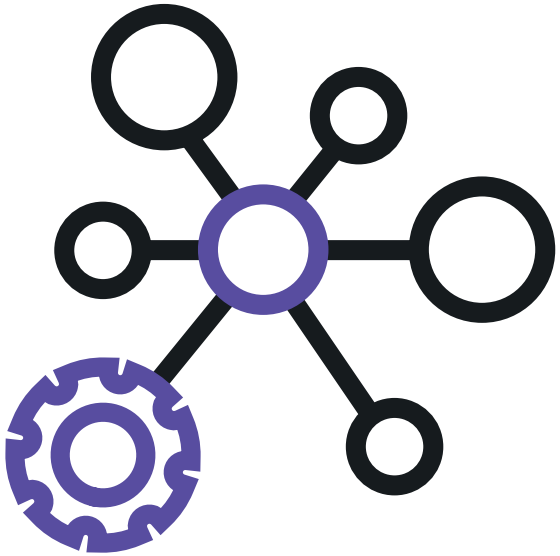
As organizations and individuals look to the future, the importance of voice technologies as part of wider digital initiatives will be vital to evolve and adapt to new ways of working and creating hybrid workforces, processes, and workflows. The pandemic has forced organizations to accelerate their automation and collaboration strategies utilizing technologies such as ASR to add flexibility to remote working while retaining important elements like data privacy and security.

The pandemic has shown that organizations that had adopted voice technology – and other future technologies – have been able to scale, pivot, adapt and operate with robustness in the face of unexpected change. These organizations have continued to thrive, while those who were not prepared or slower to adapt when the pandemic hit, were often hardest hit.

THE PROGRESSION OF VOICE TECHNOLOGY

So, what has changed in the world of speech recognition? Influential figures within speech recognition such as Dr. Tony Robinson and Professor Steve Young, pioneered the approach of applying neural networks to speech recognition in the 1980s. The approach demonstrated that neural networks greatly outperformed traditional systems. Today's computing power, along with the rise of graphics processing and cloud computing, made the huge potential of this approach a reality. The introduction of neural networks was a step-change in ASR technology.

The advent of this new approach meant that ASR became more accurate and reliable. Businesses started to realize the value that it could offer and that it could genuinely make an impact on their businesses. Speech recognition has continued to evolve to embody more than just the recognition of words in an increasing number of languages. Machine learning and more sophisticated algorithms have produced features that uplift transcription output to offer more intelligence and value to users.



In the last 20 years, voice technology has gone through a complete transformation, not only in the capability of the technology but in its accessibility and ease of adoption. In the early years, speech recognition technology had a relatively small number of uses and was available from a limited number of providers. Now, the market isn't comparable with the breadth of use cases voice technology has penetrated – both personal and professional.

Speech recognition technology is now prominent in our lives. It has come a long way from Bell Labs 'Audrey' – an automatic digit recognition machine created in 1952. This was a massive machine that occupied a six-foot-high rack, used streams of cables, and consumed a huge amount of power. Many years later – but still considered the early days – ASR providers were highly specialized and therefore there was a small competitive landscape. Speech recognition technology had limited capability which meant that there was little demand from the market and so innovation was slow.

Automatic speech recognition was prominent in popular science fiction movies and TV shows. The vision of hybrid 'man and machine' solutions and even virtual assistants were commonplace in these mediums, setting the scene for the future. Never has this vision been required or realized quite so accurately today.

WHERE IS VOICE NOW?

From humble beginnings, voice has seen a significant upward trajectory not only in adoption but also in capability. The ability to deliver quality speech recognition enables organizations to innovate with voice by leveraging any speech elements within their business (either internally, to other businesses, or the consumer). Voice is being used to add value to more use cases than ever before, from smart speakers and virtual assistants to machine interfaces in contact centers and automated ordering in drive-throughs.

Organizations are looking for more intuitive and engaging methods of interacting with customers while reducing the cost of scaling out their workforce and consumers are looking for convenience at almost any cost.

In some cases, automatic machine processes – such as ASR – can deliver equal levels of capability as humans. When it comes to transcription, it has the potential to deliver lower word error rates (WER) than human transcribers themselves. The task of transcription is not an easy one and even humans are not 100% perfect. Pioneering approaches to speech recognition leverage increasingly sophisticated machine learning and AI which has uplifted the capabilities of speech technology. It's worth considering that machine processes never get tired, they don't have bad days or struggle to focus due to other things on their minds. While errors might still occur by leveraging the latest ASR technology, organizations can greatly accelerate the speed of transcription at scale, with consistency and 24/7 availability. This means that in the same period, more content can be transcribed and then edited by humans for the final few percent.

These efficiencies not only save time but reduce cost, optimize processes, and enable additional workflows to products or services that rely on high volumes of transcription.

Organizations can now not only integrate best-in-class transcription but deliver even greater value through the continued evolution in machine learning practices that transformed speech recognition from what once was science fiction to today's reality.

Voice technology is being used by businesses for both consumer electronics through voice assistants to automate tasks and assist, and for businesses themselves looking to create efficiencies, drive down costs and deliver better customer experiences. Organizations will need to continue to focus on digital transformation, digitization, and flexibility through the investment of sophisticated technologies like voice. This will ensure they continue to drive efficiencies and profitability as the world emerges from the pandemic and market landscapes provide greater opportunities to those focused and prepared for the future.

METHODOLOGY AND DEMOGRAPHICS

To write this report, Speechmatics collated data points from Owners/Executives/C-Level, Senior Management, Middle Management, Intermediate and Entry Level professionals from a range of industries and use cases. These people work across the globe including the UK, Europe, United States, Asia and Australia.

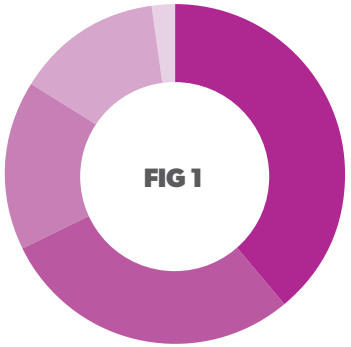


FIG 1. ROLES OF PEOPLE FROM WHICH DATA WAS COLLECTED

- **39%** Owner/Executives/C-Level
- **29%** Senior Management
- **16%** Middle Management
- **14%** Intermediate
- **2%** Entry Level

JOB TITLES INCLUDE:

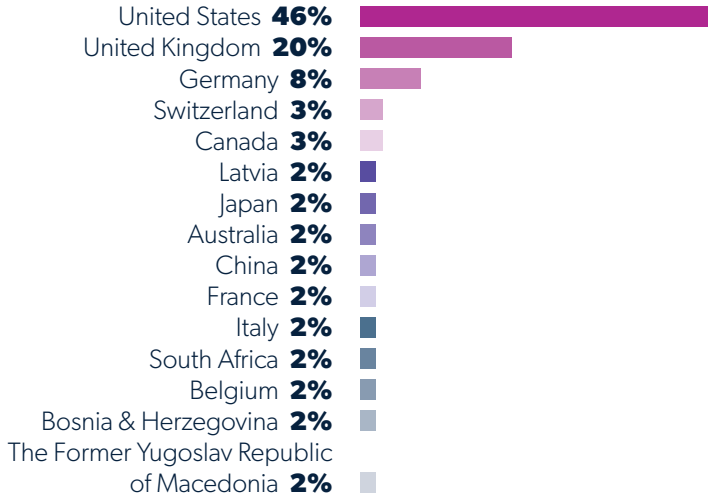
CTO
COO
CEO
Director
Managing Director
Data Scientist
Computational Linguist
Director of R&D
Enterprise Architect
Head of Archive
Head of Product
Architect Owner
Post-Production Engineer
Pre-Sales Director
Product Manager
Principal Scientist
Project Owner
Professor
Senior Software Engineer
Senior System Engineer
Speech Scientist
Task Lead
Tech Operations Director
Technical Director
Technology Advisor

The respondents described their job roles as CEO, COO, CTO, Managing Director, Director, VP Business, Director of R&D, Head of Product, Product Manager, Head of Archive, Project Owner, Senior Software Engineer, Pre-Sales Director, Post-Production Director, Conversational Consultant, amongst others. The respondent pool included a variety of organizations that operate across industries including Media and Entertainment, Technology, Transcription Solutions, Telecommunications, Banking, Financial Services, Insurance, Consultancies, Consumer Industries and Electronics, Education, Government, Healthcare, Life Sciences and Legal.

The collated data encompasses organizations from large enterprises to smaller startups. 24% of organizations surveyed employ more than 1,000 people, 3% employ 501-1,000 people, 5% employ 251-500 people, with the remaining 68% employing less than 250 people.

64% of these organizations are business-to-business, 12% are business-to-consumer, 14% a combination of the two, 5% government-to-consumer and 5% business-to-government.

FIG 2. LOCATIONS OF RESPONDENT



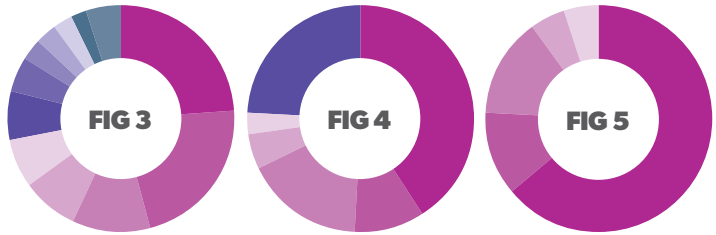


FIG 3. INDUSTRY

- **24%** Media and Entertainment
- **22%** Technology
- **11%** Transcription Solutions
- **8%** Telecommunications
- **7%** Banking, Financial Services, Insurance
- **7%** Consultancies
- **5%** Healthcare and Life Sciences
- **3%** Education
- **3%** Government
- **3%** Legal
- **2%** Consumer Industries and Electronics
- **5%** Other

FIG 4. ORGANIZATION SIZE

- **41%** 1-50
- **10%** 51-100
- **17%** 101-250
- **5%** 251-500
- **3%** 501-1,000
- **24%** More than 1,000

FIG 5. ORGANIZATION TYPE

- **64%** Business-to-business (B2B)
- **12%** Business-to-consumer (B2C)
- **14%** Mixture of B2B and B2C
- **5%** Government-to-consumer (G2C)
- **5%** Business-to-government (B2G)

CONTENT STATISTICS

Research from Deloitte as part of their their future of cloud-enabled work infrastructure research commented that "Stay-at-home orders made it difficult, if not impossible, to access on-premises infrastructure highlighting a key infrastructure risk. The vulnerability of tightly interlocked business and technology architectures to stress has become apparent. For these reasons, we expect to see a shift in cloud strategies toward cloud migration, security, operations, value planning, and DevSecOps (short for development, security, and operations) as well as a retraction of the cloud-native, container, and serverless initiatives."

In 2020, the application of speech technologies within devices like smartphones, cars, computers, and smart assistants continues to be at the forefront and drives the adoption of speech technology, particularly from a consumer perspective. Increasingly, sophisticated requests and demands from users have pushed advancements in speech technologies in consumer environments as well as long-form speech recognition. Through the development of machine learning and more sophisticated algorithms, ASR technology is continuing to deliver better accuracy than ever before.

Even research by PWC in 2018 saw that 93% of consumers were satisfied with their voice assistants. The top benefits consumers cite for using voice speakers are the ability to multitask, get instant answers to questions, and make their lives easier.

Organizations have continued to leverage the value that voice technologies can offer. Businesses utilize raw transcript outputs to fuel tools such as natural language processing (NLP), the ability to automate workflows, generate better analytics and insights, and transform speech-to-text to deliver a visual representation of dialogue as well as others.

For those who have successfully navigated 2020 and thrived by utilizing new technologies such as voice to adapt to the new ways of working, there is a huge opportunity to continue on the same trajectory into 2021 and beyond. For others who have not fared as well, the ability to leverage new technologies has never been more vital. Tools and services that help to deliver on business objectives can ensure cost effective maintenance of operations and continue to deliver the best services possible to customers.

According to Market Research Future (MRFR), the size of the global speech recognition market is expected to reach $16 billion by 2023. Market growth will remain on an upward trajectory as organizations benefit from delivering greater capabilities to their products, making interfaces more intuitive, and enable the business to optimize their processes and deliver on their goals of digital transformation.

The Adobe digital trends report 2020 found that "CX leaders are prioritizing content that's linked to the customer journey, offering more dynamic and secure experiences and engaging with customers in new arenas like voice where they have the greatest opportunity to get ahead of the market".

GOOGLE SUGGESTS THAT 27% OF THE GLOBAL ONLINE POPULATION IS USING VOICE SEARCH ON MOBILE

STATISTA FOUND THAT APPROXIMATELY 64% OF SURVEYED EXPERTS WITHIN THE INDUSTRIES OF E-LEARNING AND MARKET RESEARCH USED SPEECH-TO-TEXT AUTOMATED TRANSCRIPTION IN 2020

TECHCRUNCH REPORTS THAT THE NUMBER OF DEVICES WITH VOICE ASSISTANTS INSTALLED IN 2020 WAS OVER TWO BILLION

ROB SCOTT FROM UC TODAY CLAIMS THAT "WHEN YOU NEED TO CONVEY EMPATHY AND UNDERSTANDING TO A CUSTOMER, YOU TALK TO THEM OVER THE PHONE. WHEN YOU HAVE A COMPLEX ISSUE TO DISCUSS WITH A COLLEAGUE, YOU RELY ON YOUR VOICE. AROUND 92% OF ALL BUSINESS AND CUSTOMER INTERACTIONS STILL TAKE ADVANTAGE OF VOICE."

IN AN INTERVIEW WITH TECHCRUNCH GOOGLE CLOUD CEO THOMAS KURIAN ALSO DISCUSSED THE BENEFITS OF AI WITHIN THE CONTACT CENTER ESPECIALLY IN REGARDS TO THE PANDEMIC. BECAUSE OF THE COVID-19 PANDEMIC, MORE COMPANIES ARE NOW ACCELERATING THEIR DIGITAL TRANSFORMATION PROJECTS. KURIAN SAID THAT THIS IS ALSO TRUE FOR COMPANIES THAT WANT TO MODERNIZE THEIR CONTACT CENTERS, GIVEN THAT FOR MANY BUSINESSES, THIS HAS NOW BECOME THEIR MAIN WAY TO INTERACT WITH THEIR CUSTOMERS.

THE MICROSOFT MARKET INTELLIGENCE TEAM FOUND THAT 41% OF USERS REPORT CONCERNS AROUND TRUST, PRIVACY, AND PASSIVE LISTENING

ACCORDING TO THE ADOBE VOICE SURVEY 2020

ONE IN THREE VOICE USERS (31%) COUNT SANITATION, LIKE NOT NEEDING TO TOUCH HIGH-TRAFFIC SURFACES, AS A BENEFIT OF USING VOICE TECHNOLOGY.

86% OF USERS NOTED THAT VOICE TECHNOLOGY COULD MAKE VISITING A BUSINESS OR ATTENDING AN EVENT FEEL MORE SANITARY, VOICE TECHNOLOGY COULD BE A VITAL TOOL AS ORGANIZATIONS CONSIDER HOW TO SAFELY OPERATE IN A POST-COVID WORLD.

AT EVENTS AND BUSINESSES, HALF OF THE RESPONDENTS WANT TO SEE VOICE TECHNOLOGY FOR TASKS LIKE OPENING A DOOR (56%), CHOOSING A FLOOR ON THE ELEVATOR (55%), OR USING A VENDING MACHINE (49%).

BETTER ACCURACY IS THE MOST DESIRED IMPROVEMENT. 57% OF USERS SAY IMPROVEMENTS IN ACCURACY WOULD CAUSE THEM TO USE VOICE TECHNOLOGY MORE OFTEN OR FOR MORE PURPOSES.

KEY FINDINGS

1. Some of the current applications of voice technology include subtitling and closed-captions, customer experience and analytics, media and comms monitoring, compliance, eDiscovery, digital asset management, consumer electronics, voice bots and assistants, chat bots and medical transcription.
2. The industries that have experienced the biggest positive impact as a result of COVID-19 include healthcare and life sciences, telecommunications, media and entertainment, banking, financial services, insurance and work assistants.
3. 73% of respondents said accuracy is the biggest barrier when it comes to adopting voice technology within their business.
4. The top two voice applications that will have the largest commercial impact in 2021 are web conferencing transcription (44%) and customer experience & analytics (37%).
5. Web conferencing transcription has had the biggest positive impact as a result of COVID-19, followed by customer experience & analytics.
6. Data security and privacy, erosion of trust, a plateau in accuracy and dominance of the tech giants are the biggest risks to speech technology in the future.
7. 95% of respondents indicated that data privacy will continue to be a concern in the future.
8. Companies will overcome the challenges with data security by using on-premises deployment options and by using solutions that can be operated entirely offline.
9. 68% of respondents said their companies currently have a voice strategy.
10. 65% said that voice is likely to be considered in their business' 5-year strategy.
11. The Asia Pacific will have the largest growth in the adoption of speech recognition technology followed closely by North, Central and South America.
12. Professionals consider accuracy to mean more than word error rate. People also look at speaker change indicated, intent recognition, punctuation, number recognition and quick turnaround time for transcription when evaluating providers.
13. Automatic speech recognition is no longer synonymous with speech-to-text and additional features are expected for market to derive value.
14. 58% of respondents expect to see more robustness when it comes to transcribing voice in noisy environments.
15. 93% of respondents believe that increased demand on collaboration tools from COVID-19 will continue into 2021.
16. It is expected that model sizes will be reduced significantly, enabling more on-device deployments.
17. Less data will be required to improve language models, according to 39% of respondents.
18. People expect to see improvements in word error rate accuracy, speaker diarization, latency for real-time, language identification and customer-specific language models.

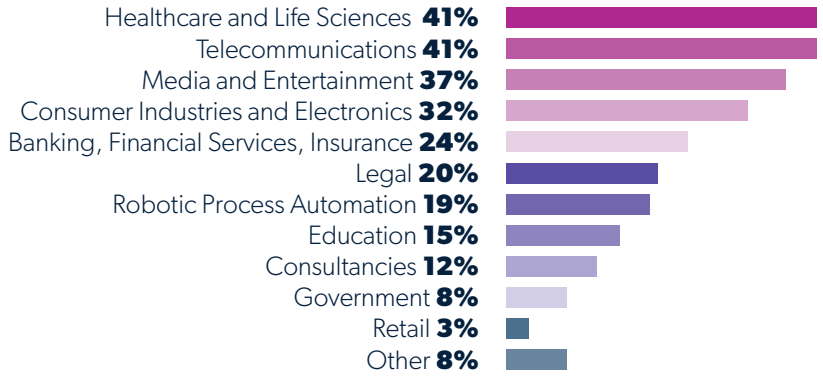
COVID-19 INFLUENCE

There is no hiding the affect COVID-19 has had on the world's economies and within those, specific industries. Some industries have experienced significant negative impact worldwide including travel, high street retail and hospitality. On the other hand, we have seen substantial positive impact on other industries as we have had to change and adapt our ways of working and living to deal with the persistent pandemic.

INDUSTRIES THAT HAVE EXPERIENCED THE BIGGEST POSITIVE IMPACT AS A RESULT OF COVID-19

Respondents think that the industries that have experienced the biggest positive impact as a result of COVID-19 are:

FIG 6.



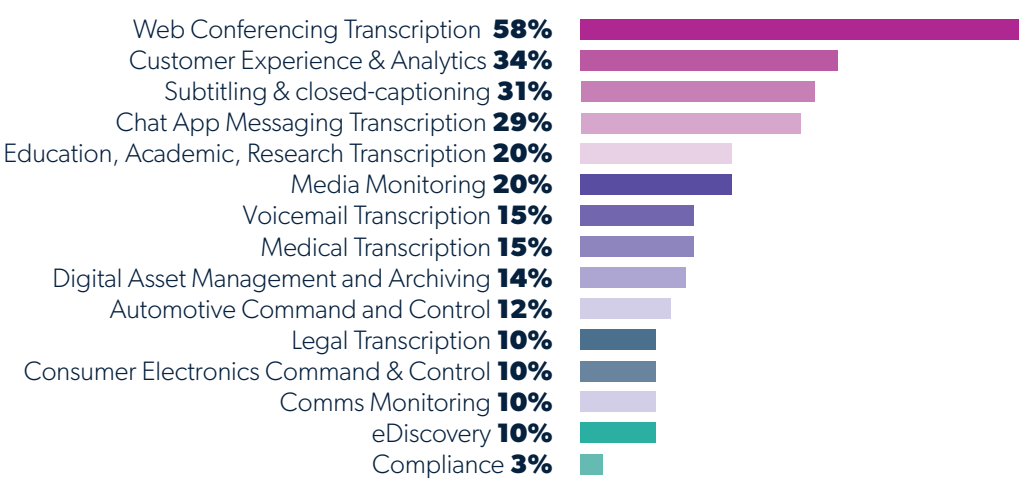
41% OF RESPONDENTS THINK THE HEALTHCARE AND LIFE SCIENCES AND TELECOMS INDUSTRIES have experienced the biggest positive impact as a result of COVID-19 in 2020. The healthcare and life sciences industry has been in the spotlight around the world during the pandemic. Not only continuing to develop healthcare breakthroughs generally but also the development of new vaccinations and ventilators. The telecoms industry is also seen to have experienced a positive impact in 2020 due to the increased volume of calls in an ever increasingly connected world. Due to the pandemic, we now have to almost solely rely on telecoms to work and socialize so it's no surprise this industry is seen to be growing at such a rapid rate.

37% OF RESPONDENTS THINK THE MEDIA AND ENTERTAINMENT INDUSTRY has also experienced a significant positive impact as a result of COVID-19. The growth in demand for social media, online videos and over the top (OTT) streaming services since the beginning of 2020 has been substantial and this is only expected to continue on the same trajectory into 2021 and beyond.

USE CASES THAT WILL HAVE EXPERIENCED THE BIGGEST POSITIVE IMPACT AS A RESULT OF COVID-19.

Respondents think that the use cases that will have experienced the biggest positive impact as a result of COVID-19 are:

FIG 7.



With the overwhelming majority, **58% of respondents think web conferencing transcription will have experienced the biggest positive impact as a result of COVID-19**. With web conferencing platforms being adopted to suit home working across the world, the subject of captions and transcription has cropped up continuously. Web conferencing has become the best – and in some cases only – way of communication overnight. Savvy businesses recognized a huge opportunity when it came to accessibility of these web conferences. The requirement for captions became a must to ensure conversations were accessible and understood by everyone. Transcripts soon become equally as important as a way of capturing the conversations as they happen so businesses could capture, store and analyze what is said on calls – think of this as automatic meeting minutes. Many industries have since turned to web conferencing and with that comes many challenges without transcription. Transcription enables web conferencing calls to be transformed into text which can then be analyzed, reviewed and stored using metadata, providing a 360-degree view of any interaction or used simply as meeting minutes.

34% of respondents think customer experience & analytics has seen a significant positive impact as a result of COVID-19. Since the pandemic became an international emergency, customer experience has been highlighted as a major issue within contact centers. Not only are agents now dealing with more customer issues by phone, but customers are also more vulnerable than ever and so must be treated sensitively on a case-by-case basis.

HAS COVID-19 HAD PROFOUND EFFECT ON THE MASS ADOPTION OF VOICE TECHNOLOGY INTO BUSINESS AUTOMATION WORKFLOWS?

53% of respondents believe COVID-19 hasn't had a profound effect on the mass adoption of voice technology into business automation workflows, while 47% think it has. Looking at the data in this report as a whole, many respondents already had a voice technology strategy implemented and regard voice as an already widely adopted technology. This could be an explanation for the even split.

The respondents that believe COVID-19 has had a profound effect, recognize that many industries and use cases had significant gaps in their voice technology strategies when the pandemic hit. Businesses in these industries – such as web conferencing – assessed the risks of not adopting a voice strategy and opportunities with adding voice technology into their workflows and have seen significant growth as a result of wide adoption.

WILL THE INCREASED DEMAND ON COLLABORATION TOOLS FROM COVID-19 CONTINUE INTO 2021 AND BEYOND?

The overwhelming majority of **respondents (93%) think that the demand on collaboration tools will increase into 2021 and beyond.** 2020 has set a new way of working and living across the world. It has taught businesses that many jobs can be done from the comfort of home with no need for commuting or expensive office leases. This new approach to working – whether that is a hybrid home/office approach or full remote working – has now become the norm and many businesses have stated that they wouldn't return to the way they were operating before the pandemic hit. As a result, businesses will become more reliant on collaboration tools to ensure business stays efficient and productive.

The increased demand on collaboration tools has also enabled businesses to rethink their recruitment strategies. With this new way of working – that is proving for most to be effective from a business and employee wellbeing perspective – businesses can look beyond their location-based talent pools and recruit from anywhere in the world.

2% of respondents selected 'other'. One respondent explained that it was impossible to say either way since the future is not predictable – a pertinent point in these uncertain times.

> **"The real change we have seen in 2020 is the world adapting to COVID-19 and with that, the increased uptake of contactless engagement systems to avoid spreading germs. Speech has a big play here and might be impacting markets. This also reflects the large growth seen in the chat bot markets as contact centers look to scale up due to increased inbound communications for what are becoming FAQs that are related to the businesses.**
>
> **Since COVID-19 is not going away anytime soon this trend is set to continue, and I am not sure we will ever go back to where we were before as people get used to a new way of working."**
>
> **Thuy Le**, Senior Product Manager, Speechmatics

OVERVIEW OF VOICE TECHNOLOGY USE CASES

PROFESSIONALS FROM A RANGE OF INDUSTRIES WERE SURVEYED, INCLUDING:

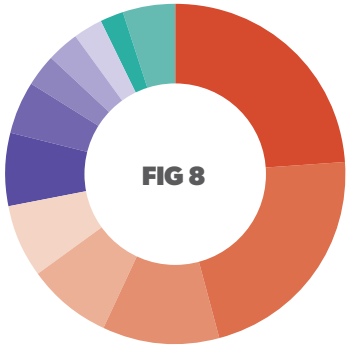


FIG 8. INDUSTRY

24%	Media and Entertainment
22%	Technology
11%	Transcription Solutions
8%	Telecommunications
7%	Banking, Financial Services, Insurance
7%	Consultancies
5%	Healthcare and Life Sciences
3%	Education
3%	Government
3%	Legal
2%	Consumer Industries and Electronics
5%	Other

Voice technology has become widely adopted to support businesses with their digital transformation projects. The popularity of the technology has been led primarily by the consumer adoption of digital voice assistants like Alexa, Siri and Google Home. As with most consumer-led technologies, voice technology has matured, and businesses are turning to its capabilities to improve efficiencies and revenues. Some industries that voice has benefitted in the last year include media and entertainment, banking, financial services, insurance and transcription solutions.

MEDIA AND ENTERTAINMENT

The rise in production of video and audio content, along with the creation of new channels means that media companies have been challenged with managing and monitoring what is said within this data. Media and broadcast companies are using automatic speech recognition technology to get ahead of their competition.

The technology is being used in media monitoring to set live triggers on chosen keywords. It is also used in digital and media asset management to transform video and audio files into searchable and indexed transcripts filled with enriched metadata. It can also be used to form real-time or pre-recorded captioning for use in broadcast scenarios and for better accessibility of video content.

Some key drivers and motivations for media companies adopting voice technology include reduced costs, productivity improvements, support/assistance for human tasks, developing better insights, operational efficiencies, improving accessibility of content, generating competitive advantages and product enhancements.

BANKING, FINANCIAL SERVICES AND INSURANCE

The global pandemic coupled with rising consumer expectations has led to a huge rise in demand for contact centers. This increase in call volume has put pressure on the banking, financial services and insurance industry to adapt and ensure all calls are monitored and recorded for compliance purposes. With regulators putting more pressure on this industry to remain compliant and ensure the data security of all customer calls, contact centers are using automatic speech recognition to transform customer calls into text which can be easily managed and stored. With all calls available as transcripts, contact centers can locate and replay stored recordings for a range of applications including creating best practices, ensuring compliance, quality management and event reconstruction.

As well as compliance, the financial services industry is using automatic speech recognition technology to improve customer experience through voice data analytics and other strategies. Banks and other financial services organizations are using customer interactions with contact centers to provide unrivalled customer experience using speech recognition. Customer calls can be transformed into valuable insights to help with practices such as issue resolution and providing an agent knowledge base. Turning the customer voice into text enables contact centers to analyze their call content and understand the mood, tone and overall sentiment of customers. This supports continuous improvements in customer experience.

Some key drivers and motivations for contact centers adopting voice technology include, improve the productivity of support staff, increase customer satisfaction, reduce staffing needs, improve agent related KPIs and knowledge, unlock insights to empower and drive change.

TRANSCRIPTION SOLUTIONS

Video and audio content is being created and distributed across more channels than ever before. Transcription companies are providing a service to transform these video and audio files into text for users to benefit from better SEO, captioning capabilities and improved accessibility of their content.

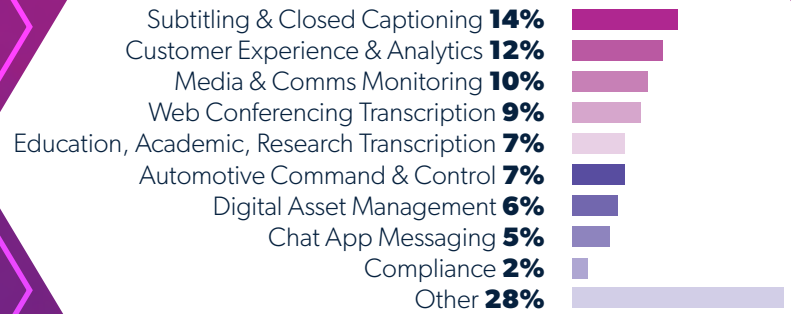
Companies are providing transcription services to generate searchable, editable transcripts for their customers quicker and more accurately than ever before. Features such as speaker identification, highlight and comment functionality, adjustable timestamps and a custom dictionary make this process streamlined and efficient. The technology enhances the speed and accuracy at which transcripts are created, taking the heavy lifting away from manual transcribers and enabling them to add value where only humans can.

Some key drivers and motivations for transcription solutions adopting voice technology include, take heavy lifting and mundane tasks away from manual workers, improve efficiency, save time, reduce cost and speed up their service.

CURRENT USE CASES FOR VOICE TECHNOLOGY

RESPONDENTS SURVEYED THINK THE MAIN CURRENT USE CASES FOR VOICE TECHNOLOGY INCLUDE:

FIG 9.



SUBTITLING & CLOSED CAPTIONING

In 2018, the global Captioning and Subtitling Solutions market size was 220 million USD and it is expected to reach 370 million USD by the end of 2025, according to Valuates Reports.com.

Captioning and subtitling solutions enable the transformation of video and audio content in to a text-based format that can then be used to deliver captions automatically, quickly and at scale. The automation of captions and subtitles provides human transcribers/editors with the capability to review and simply tweak the output. Captioning solutions can also aid in the editing, encoding and repurposing of video subtitles and captions which can then be published or broadcasted on a wide array of applications and hardware devices.

The use of ASR for captioning and subtitling solutions enables broadcasting and media organizations to process high volumes of captions faster and cheaper than ever before.

Captioning is a highly legislated practice across the world and especially in the US where the Federal Communications Commission (FCC) regulates online and broadcast captioning and accessibility. For this reason, it's vital that captions are 100% accurate no matter whether the content is available from a linear service (broadcast TV) or from an online source (an OTT streaming site, social media platform or other online source).

The pandemic has had a significant impact on many businesses that have been forced to market their products and services differently. With stores on the highstreets closing and the cancellation of face-to-face marketing opportunities, brands have doubled down on online advertising and marketing. Captioning has a significant part to play for digital marketing due to the additional engagement and need for accessible media content.

2020 saw a dramatic increase in the volume of video content created and consumed compared to recent years.

- Global content delivery network provider CDN Akamai reports that global internet traffic has grown by as much as 30% this year
- July 2020 saw a rise of 10.5% in social media usage, compared with July 2019, according to a GlobalWebIndex survey
- In their study of how the Coronavirus pandemic has been influencing people's digital behaviors, GlobalWebIndex has found that more than 40% of internet users have been spending more time using social media in recent months
- The State of Video Marketing report 2020 by Wyzowl claims 92% of marketers who use video say that it's an important part of their marketing strategy
- Since early March of 2020, video marketing software provider, Wistia, saw a year-on -year increase of 120% average number of hours of video content consumed per week across all customers. A drastic increase from 2.6 million in 2019 to 4.6 Million in the same period of 2020
- According to Hubspot, 85% of businesses use video as a marketing tool
- A study by leading social media platform Facebook claims that people now watch over 100 million hours of video on Facebook each day. Additionally, they said that internal tests showed that captioned video ads increase video view time by an average of 12%

A study by leading social media platform Facebook claims that people now watch over 100 million hours of video on Facebook each day. Additionally, they said that internal tests showed that captioned video ads increase video view time by an average of 12%.

From our research, we found that **14% of respondents called out that subtitling & closed captioning** is the main application for speech technology, putting it at the top of the list at the end of 2020.

The benefits of automatic speech recognition further support this trend of increased content creation. ASR offers brands, editors, and creators the tools to caption content with speed and quality through the use of automatic transcription tools included within their editing and creation suites. The ease of use of transcription as part of these platforms not only makes it easy to navigate the content as it is being created but also through the ability to deliver additional value like captioning to the finished product. Here, high-quality automatic speech recognition wins out through the ability to deliver quality and convenience even as the volume and demand for content continues to increase.

With the pandemic continuing and with the 'new normal' further establishing itself it's likely that this trend of video content creation will continue to offset other marketing strategies that can no longer be used like face-to-face events and physical retail stores. Captioning and subtitling will continue to be a required application of speech technology not only for high-quality broadcast content but for all content to aid in accessibility no matter the audience.

CUSTOMER EXPERIENCE & ANALYTICS

Grand View Research found that the global contact center analytics market size was valued at over $930 million in 2019 and is expected to expand at a compound annual growth rate (CAGR) of 16.8% from 2020 to 2026.

Contact centers have been at the frontline of adopting new developments when it comes to voice technology. As an industry and market where voice is the primary source of interaction and data, contact centers are motivated to leverage continued advancements in voice technologies to optimize their businesses. By capturing, structuring, and analyzing data derived from voice they can understand patterns in data and even predict future outcomes. Attitude to voice has continued to evolve in line with its capabilities as consumer expectations have changed and legislation has brought in new rules that must be respected for the protection of consumer data. Contact centers have been able to adapt to these changes through the use of voice technologies and have provided insight and influence on how these technologies need to evolve to meet their demands.

Originally, call recording was enough for contact centers to keep track of interactions and ensure compliance. While this provided a solution in the short term, recordings of audio conversations are hard to index, search, and interrogate especially when the calls need to be investigated quickly in a dispute situation. Calls in audio format also require significant storage space. With customer experience being a core factor of measurement in the contact center and unhappy customers resulting in loss of revenue and loyalty for the represented organization, analysis needs to quick and cost-effective.

The ability for contact centers to transcribe their content provides many advantages. Interactions have become easier to index and search if they need to be found quickly. Agents are empowered to significantly reduce the time taken to resolve disputes and the contact center can innovate their solutions by transforming the audio from calls to a text-based format. When in text, call recordings can be added into natural language processing tools that already exist in the contact center to gain insight from omnichannel approaches like text bots, instant messaging, and email interactions with customers. The archives of existing call recordings in contact centers are a potential gold mine of data that voice technology can transform into key insights.

In 2020 and continuing into 2021, contact centers have been challenged to meet unprecedented demand from callers and, in some cases, with teams that are stretched due to illness. Contact centers have had to entirely redefine their working practices. Where previously, contact center staff were in a single working environment, due to social distancing and working from home orders, they have been forced to change their operation to support their staff who are now – mostly – required to work remotely.

COVID-19 research found significant volumes in calls to contact centers often related to the same or similar topics. This is an ideal use case to leverage analytics where commonality can be found and resolution tactics applied – such as online FAQs – to reduce calls, call durations and increase the overall customer experience.

"ASR systems are becoming more prevalent in contact center solutions. Historically EU-based financial services have been reluctant to engage with ASR due to onerous data localization and restrictions, but with on-premises solutions that comply with regulations like GDPR, more are experimenting with solutions. I expect this trend to accelerate as contact centers and customer services continue to become the virtual front door of an organization."

Michael Tansini, Product Owner, Speechmatics

WEB CONFERENCING TRANSCRIPTION

Web conferencing was already a growing industry before the pandemic with a vast number of businesses already utilizing the likes of Zoom, Teams, and Webex amongst others. Some businesses were well versed in collaboration tools and already set up, others had to adapt and learn quickly.

While many providers were available, Zoom adapted far more quickly to the pandemic, seeing the opportunity of its service for things like education and connectivity for those unfamiliar with collaboration who could easily adopt its service through its ease of use, and importantly, zero price tag.

In 2017, CEO of Zoom, Eric Yuan said: "our philosophy is we focus on making our existing customers happy. We do not aggressively pursue the new prospect."

Zoom took a customer-focused approach, and ensured user experience was optimized ensuring mass adoption – even if this did cause several problems later in 2020 including server capability and privacy.

Voice technology as part of tools like Zoom, Teams and Webex has continued to evolve and for the most part, these platforms already have voice capabilities like speech-to-text. This means that they can transcribe calls as they happen or post-call, depending on the service or the option chosen by the user.

part of web conferencing tools is a standard feature. Any new organization looking to take a piece out of this extensive and lucrative market in the wake of the pandemic needs world-class transcription as a minimum in this rapidly evolving market.

For businesses, increased reliance on these tools means that communication with their teams can be maintained and the accessibility of these communications can be achieved through real-time captioning and the ability to create a full transcript once interactions cease. Additionally, organizations now have access to the interactions of their staff through the ability to monitor and record calls which increasingly occurs on platforms (internal and external ones) managed and provided by the business. Where previously meetings might happen face to face, over a water cooler, or via a call, web conferencing means that interactions moved into an environment that delivers better control and intelligence to the business.

EDUCATION, ACADEMIC, RESEARCH TRANSCRIPTION

Web conferencing tools also have a significant impact on education as this became the virtual portal into the physical classroom in 2020. With schools shut, video conferencing technology enabled millions of people (children and adults) to remain in education. Video conferencing technology gave access to lessons from home at an unprecedented scale.

Global market insights **predict that the e-learning market** size surpassed USD 200 billion in 2019 and is anticipated to grow at over 8% CAGR between 2020 and 2026 to 375 billion USD by this time.

In April of 2020, the world Economic forum commented that while countries were at different points in their COVID-19 infection rates, at that time, worldwide there were more than 1.2 billion children in 186 countries affected by school closures due to the pandemic.

Even before COVID-19, there was already high growth and adoption in education technology, with global educational technology investments reaching US $18.66 billion in 2019. The overall market for online education is projected to reach $350 billion by 2025. Whether delivered through language apps, virtual tutoring, video conferencing tools, or online learning software, there has been a significant surge in usage of e-learning platforms since COVID-19 to ensure that education is maintained no matter the challenges set by the pandemic.

Online education platform, Tencent, reported that at noon on February 10th, 2020 730,000 primary and middle school students in Wuhan, chose to use its service for online live learning.

This market will likely continue to see a significant uptake as lockdowns and restricted movements continue to impact people in education. While tools exist to connect teachers and students, voice technology plays a continuing part in the services and tools used to deliver education. Captioning – much like how is used through web conferencing – ensures that interactions/lessons can be understood in more than just a verbal medium.

The means to have lessons transcribed in real time helps people track what is being said, whether hard of hearing or not. Additionally, the capability to download a transcript at the end of a lesson provides an additional learning tool to help extract as much value as possible from virtual interactions.

Alibaba's distance learning solution, DingTalk, which services more than 10 million businesses and organizations globally, stepped up measures to help schools continue their lessons during the Coronavirus outbreak. In January 2020 – shortly after China postponed the start of the new school semester – the app launched an Online Classroom initiative to provide schools with free digital tools, such as live streaming and online testing and grading features. About 120 million students and 140,000 schools across the country were able to resume classes through the app.

Transcription for education use cases presents a unique challenge due to the extensive domain-specific language used in educational classes. It remains a major challenge to ensure accuracy and the usability of transcripts as educational use cases continually evolve and adapt.

AUTOMOTIVE COMMAND AND CONTROL

Voice assistants has been a growing trend in recent years with the majority of technology providers looking to incorporate this technology into existing and new products to enable users to engage via voice. With increased scrutiny and concern for hygiene since the pandemic began, voice is becoming a more attractive interface for products that have until now, required human touch to operate.

It is expected that a continued drive toward technologically advanced passenger vehicles with an increased level of sophistication, capabilities, gadgets, and features will compete with the safety aspects of controlling these functions. The application of voice technology has the potential to remove menu trees which are becoming increasingly complex in modern passenger and commercial vehicles. For this reason, voice interfaces make interaction in vehicles safer on the move.

According to Acumen Research and Consulting, the global automotive voice recognition system market is expected to reach the market value of around US $39 billion by 2027 and is anticipated to grow at a CAGR of around 21% in terms of revenue during the forecast period 2020 – 2027.

Automotive applications of voice technology enable drivers to control their surroundings to limit distracting factors. Command and control functions like those used with voice assistance like Alexa are finding themselves within automotive environments to aid drivers in controlling both elements within their car and outside of it. From satellite navigation to turning up the volume of the stereo or interacting with a mobile device through the infotainment system built into the car. No matter the action or the command delivered, at the root of the workflow, speech needs to be transformed into text. This text-based output then powers all other elements in the workflow so it's important to get right.

Research by Capgemini found that out of the top 100 organizations surveyed as part of their research into the technology used in the automotive industry, 48% of automotive organizations are deploying voice assistants within their vehicles. In the same report, Henrik Green, Senior Vice President of Research and Development at Volvo Cars commented that "Soon, Volvo drivers will have direct access to thousands of in-car apps that make daily life easier and the connected in-car experience more enjoyable."

Guardian reported on March 24th 2020, that, in the film industry, the US box office had recorded zero revenue for the first time in history after the country's near-total cinema shut down in response to the Coronavirus pandemic.

In reaction to limited revenue generated through traditional cinema ticket sales, 2021 will be the first year in history where high-value content like 'box office' movies will be available to view in the home in parallel to theatres. In an interview between Pav Kudlac, MD and Co-Founder at Ovyo, and Steve Huin, CMO at Irdeto, they predict that there will be a broad adoption in watermarking or it will be a big year for piracy.

The huge volume of content created and available at the end of 2020 and leading into 2021 means that organizations will need laser focus on content curation to manage their assets. Consumers will require additional tools and capabilities to help them navigate the plethora of content available to them. Organizations that deliver streaming solutions, content producers and creators have the opportunity to leverage automation tools like speech-to-text and transcription to unlock the keywords, themes, and other elements contained within their media content to further help in its consumption and accessibility.

Delivering the means to unlock greater levels of insight across a large volume of content automatically, unlock metadata, and use the data to help users find the content they enjoy significantly optimizes their experience. The ability to unlock metadata to offer personalized content further optimizes the viewing experience and drives engagement through recommendation. These

solutions are built on an understanding of elements that can be extracted through speech-to-text and other machine learning and AI-derived tools. For brands, this provides an opportunity to optimize customer retention and revenues.

Matthew Eaton, Managing Director, EMEA predicts that "in 2021 content owners will need to innovate more around their use of data. Having an understanding of the content, metadata will be a key component in storytelling". The use of tools and technology to extract metadata will aid content producers in understanding their content, how and where to record it, and importantly evaluate its performance – what remains on-trend and what needs to be replaced to ensure best ROI.

Asset management remains vital for editing and efficiency of finding the right content. The speed of creation will likely remain important, however, finding the right element within a specific piece of content can be challenging. It's expected that voice technology and other tools able to expose key metadata from media content will be in equal or greater demand as media content remains a vital medium in 2021.

CURRENT MARKET ADOPTION OF VOICE TECHNOLOGY

DOES YOUR COMPANY HAVE A VOICE STRATEGY?

68% of respondents said that their company currently has a voice strategy, up 18% on last year.

This year has seen an increase of 18% in businesses that have adopted a voice strategy. 68% of respondents indicated that they have a voice strategy compared with 50% last year. Of those that don't have a voice strategy, 60% said it is something that will be considered in the next 5 years.

Voice technology is becoming a core component of many business's technology stack and strategy. It unlocks many other capabilities only possible by transforming speech into text. Voice technology currently sits between the early adopters and early majority in relation to the product adoption lifecycle. Voice is the easiest form of communication, and we predict that there will continue to be a major shift in communication within enterprises in 2021. Over the next 5 years businesses will start to fully integrate voice technology into their workflows.

CURRENT APPLICATIONS FOR VOICE TECHNOLOGY

Respondents indicated that there are currently many applications for voice technology. They are listed below

- Subtitling & Closed Captioning
- Customer Experience & Analytics
- Media & Comms Monitoring
- Web Conferencing Transcription
- Education, Academic, Research Transcription
- Automotive Command & Control
- Chat App Messaging
- Digital Asset Management
- Compliance
- eDiscovery
- Voice Bots & Assistants, Chatbots
- Accessibility
- Call Center Transcription
- Voicemail Transcription
- Medical Transcription
- Consumer Electronics
- Interview Transcription

While last year lots of the applications for voice technology were consumer-focused, this year respondents have indicated more business applications for the technology. Consumer voice applications have put voice technology on the map over the past few years. The likes of Siri, Alexa and Google Home brought the possibilities of voice communication to millions of devices. However, voice technology in consumer devices is currently restrictive and only enabled via short utterances or commands. In contrast, business applications are thriving.

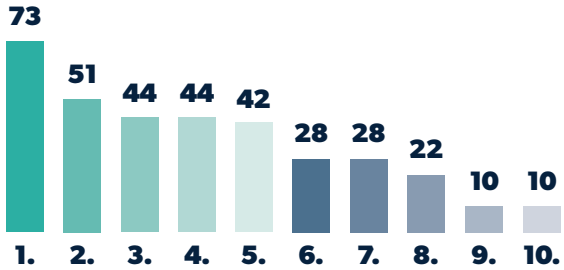
From accessibility related applications such as closed-captioning and subtitling, to digital and media asset management, compliance monitoring and speech analytics, voice technology brings operational efficiencies and improvements to businesses of all sizes in any industry.

BARRIERS TO ADOPTING VOICE TECHNOLOGY

Barriers to entry are commonly the reason why businesses are late to adopt innovative technology. When it comes to voice technology, it has been a common misconception for many years that the technology isn't good enough to adopt as an integral part of a workflow and technology stack. This is simply not true anymore. Over the past few years, voice technology has improved to a point in which the output for the most spoken languages in the world such as English, French, Spanish and German is highly accurate in terms of word error rate. It's at this stage where other challenges and factors affect the rate of adoption.

As businesses have started to overcome this misconception and consider voice technology in their 5-year innovation and technology strategies, what other barriers to adoption are proving challenging? And why is accuracy still a problem?

FIG 10. SOME OF THE BARRIERS TO ADOPTION OF VOICE TECHNOLOGY ARE INCLUDED BELOW.



1. 73% Accuracy
2. 51% Accent or dialect related recognition issues
3. 44% Language coverage
4. 44% End-user expectation
5. 42% Data security and privacy
6. 28% Cost
7. 28% Complexity of deployment and integration
8. 22% Flexibility to meet the use case
9. 10% Operational support
10. 10% Other
    - Lack of awareness on its benefits
    - Sophistication to derive meaning from voice
    - Backlash from human transcription professionals

ACCURACY

73% of respondents believe that accuracy is the biggest barrier when it comes to adopting voice technology within their business. These days, accuracy represents more than just the accuracy of the word output – known of as word error rate (WER). With the most spoken languages in the world at a consistently low WER, many other factors affect the level of accuracy on a case-by-case basis. These factors are often unique to a use case or a business' needs.

DATA SECURITY AND PRIVACY

Data security and privacy is said to be one of the largest barriers to adopting voice technology with 42% of respondents indicating this. To put this into perspective – last year, just 5% of respondents perceived data security as a barrier to adoption. We believe this dramatic increase could be down to the mistrust in the market brought about by the media portrayal of the 'data-hungry' tech giants. It could also be a result of more day-to-day conversations happening online with the world being thrown into remote working overnight.

DEPLOYMENT

Deploying and integrating voice technology – or any software for that matter – needs to be simple. **28% of respondents said the complexity of deploying and integrating voice technology is a barrier to adoption**. Whether a business requires deployment on-premises or in the cloud, integration needs to be easy to do and secure. Without the appropriate support or documentation, integrating voice technology can be time-consuming and expensive. It is, therefore, important for technology providers to make their deployments and integrations as frictionless as possible to avoid this barrier to adoption.

FACTORS AFFECTING THE LEVEL OF ACCURACY INCLUDE:

- BACKGROUND NOISE
- PUNCTUATION PLACEMENT
- CAPITALIZATION
- CORRECT FORMATTING
- TIMING OF WORDS
- DOMAIN-SPECIFIC TERMINOLOGY
- SPEAKER IDENTIFICATION

LANGUAGE COVERAGE

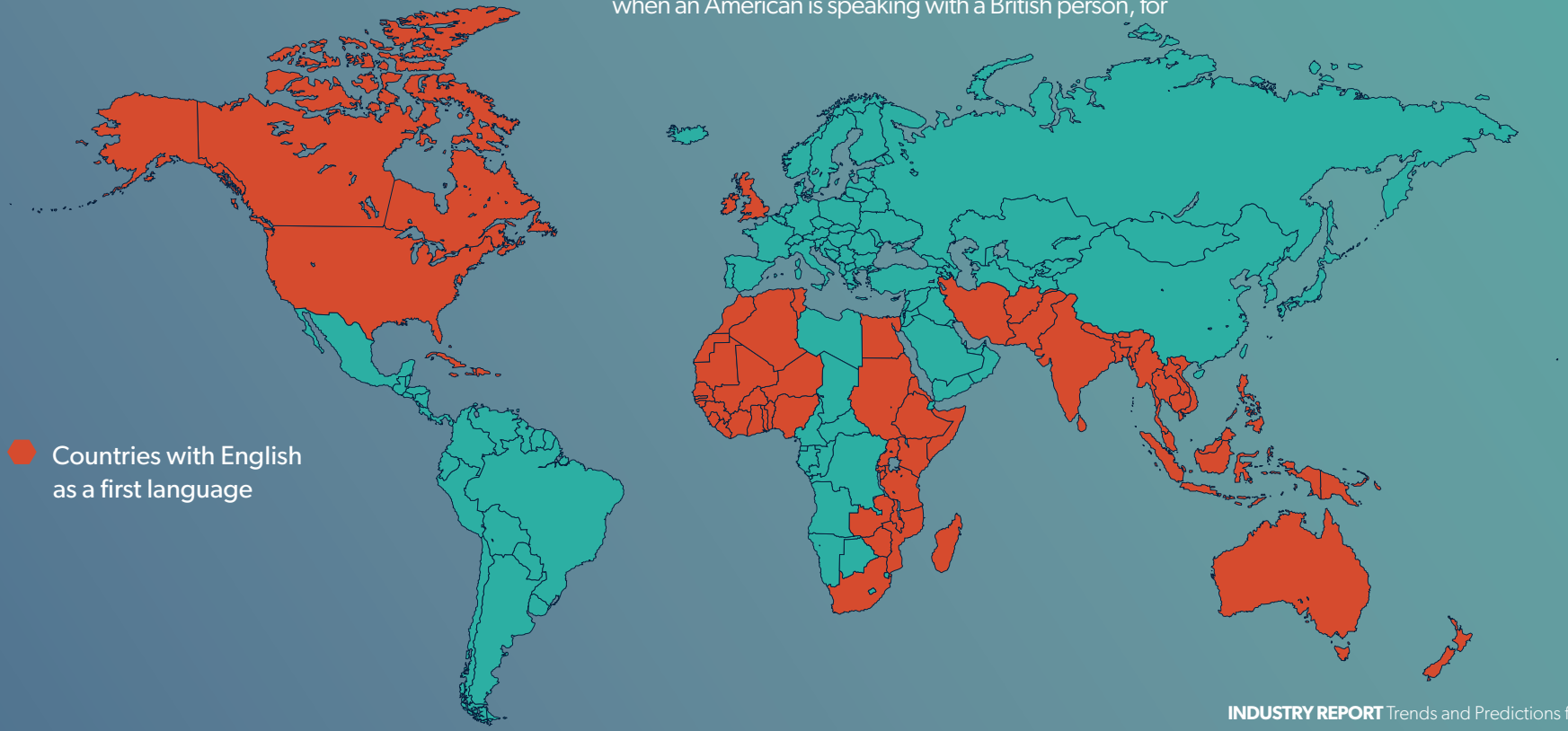
44% of respondents identified language coverage as a key barrier to entry. Many of the leading voice technology providers have a gap when it comes to language coverage. Most providers cover English but when global businesses want to use voice technology, the lack of language coverage provides a barrier to adoption.

When providers offer more languages, an important question to ask is around the quality of output.

How accurate are those additional languages? The development of a range of accurate languages will help businesses to adopt voice technology into their workflows.

On top of language coverage, **51% of respondents labeled accent or dialect related recognition issues** as a barrier. Yes, a provider may offer English, but which accents and dialects do they support? Providers are often challenged by thick accents, causing issues when it comes to the transcription output. What happens when an American is speaking with a British person, for example? Which accent variation is used? If you choose American-English it will be at the detriment of the British speaker, and vice versa. Organizations ordinarily have to run two transcripts using both language packs. This is time-consuming and expensive.

A global approach to languages solves this challenge. By incorporating data drawn from global sources, encompassing a variety of accents, providers can offer global language packs, so users don't have to worry about different accents in their video or audio.



Countries with English as a first language

FUTURE APPLICATIONS OF VOICE

The future applications of voice technology are potentially limitless. 2020 has demonstrated the need to expect the unexpected, and on a global scale. The challenges delivered in 2020 have been met by some seriously impressive innovation. The wide adoption of technologies like web conferencing have been enabled by a development in fast connectivity, delivering low latency interactions, dynamic and elastic cloud resource scaling, and can now support millions of concurrent users. Machine learning and AI is being used to deliver additional value to these platforms – like speech-to-text – to optimize the user experience.

CLOUD TECHNOLOGIES

Research from Deloitte found those firms that fared better in the pandemic had already adopted virtualization and cloud technologies. Those that hadn't invested scrambled to do so; PwC reports spending on cloud rose 37% during the first quarter of 2020.

The research predicted that 87% of global IT decision-makers agree the pandemic will cause organizations to accelerate their migration to the cloud, anticipating a decline in on-premises workloads by 2025. With this being said, it's expected that data security is still a major concern for both businesses and individuals when using cloud services.

IMPROVED CUSTOMER EXPERIENCES

Global businesses looking to deliver better services to their customers continue to push the evolution of voice technologies and make them accessible and useful to global audiences. This means voice providers must support a greater diversity of languages and offer better accuracy across their entire language offering.

Voice continues to be a genuine and viable interface for interacting. Demand is also increasing with businesses forced into limiting physical interactions. Consumers are looking to voice to simplify their lives and enable additional multitasking in a hands and eyes-free manner.

Consumers are also becoming more trusting of voice technology within business applications. Voice-based IVRs are becoming more common and have delivered significant advantages and support to contact centers who have faced high caller demand and staff shortages due to the pandemic. Better transcription quality fosters better responses and enhanced customer experiences. These better experiences drive adoption, enabling businesses to maintain the trust of their customers while deploying solutions that improve ROI and drive down costs.

Organizations recognize the importance of harnessing their data, whatever form it may present itself in. Due to the rise in generation of audio and video assets, it's expected that voice technology solutions will continue to be in demand to surface more data trapped in video and audio files. Voice technology will help businesses to better manage, evaluate and analyze this data, along with other data already accessible in text form such as customer emails, text, chat bots etc.

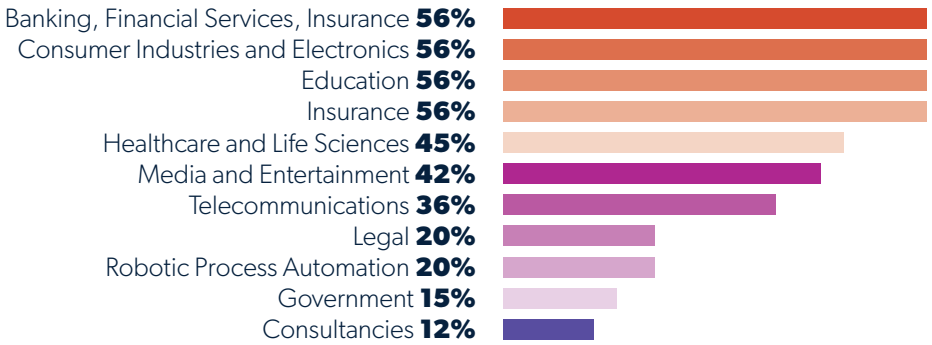
AI BIAS

2020 included more social and economic events than just COVID-19. These events had an impact on the potential future use and development of voice technology. Bias continues to be a heavily discussed topic within the AI and machine learning fields. This then has a direct impact on tools like speech recognition that are developed through these processes.

It is expected that innovation will remain strong in 2021 with the development of new languages and features. In addition, extensive effort will be directed at continuing to diversify and optimize existing languages and product capabilities. Even popular and mature ASR languages require retrospective improvements to ensure they represent as many users as possible through the support of more accents and dialects. Development and evaluation of existing and new language to remove bias will ensure better performance and accuracy of all tools and solutions that leverage ASR as part of their workflow.

INDUSTRIES THAT WILL INCREASE THEIR USE AND APPLICATION OF VOICE TECHNOLOGY IN THE NEXT 3-5 YEARS.

FIG 11. RESPONDENTS THINK THAT THE FUTURE OF VOICE TECHNOLOGY WILL INCLUDE:



56% of respondents believe that the banking, financial services and insurance industry will increase their use and application of voice technology in the next 3-5 years. Capgemini predicts that over the next three years, 70% of consumers, on average, will replace their visits to the dealer, store, or bank with their voice assistants. Accuracy of speech-to-text services has increased to a level where users are now confident using them – even for sensitive transactions like banking.

Additionally, convenience continues to be a massive driver for consumers. It's predicted that this convenience will continue to overshadow what could be perceived as issues around the security of using voice technology from cloud services for banking and finance uses. Customers can be inclined to overlook potential security issues if the ability to use voice to access services is faster and more convenient than before.

This change in consumer behavior puts additional pressure on speech recognition providers to deliver effective number recognition in all of the ways individuals might say a number, for example naught, zero, nill, oh and ensure that this can be understood by the interface that they are using.

56% of respondents said that consumer industries and electronics will increase their use and application of voice technology in the next 3-5 years. Faster connectivity means that real-time voice solutions are no longer confined to areas where broadband WiFi is available. The appetite for voice services continues to grow and recently through the roll out of 4G and 5G, the tools are available to support it. From automotive applications to virtual assistants on phones, the technology exists to bring voice services to the user no matter where they are. The combination of low latency, real-time voice services with the optimized UX of a mobile phone enables this at scale.

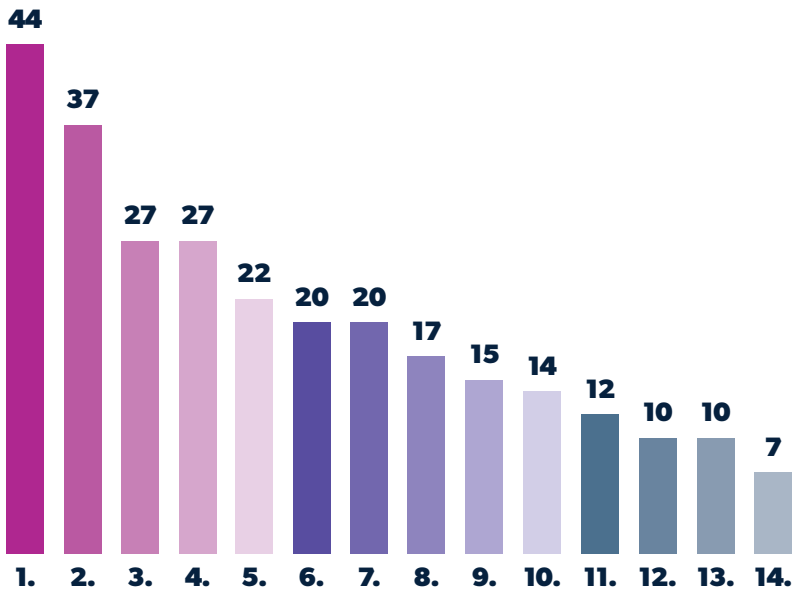
Cloud-based voice services are not the only route for delivering voice services to consumer devices. However, as most mobile phones currently already operate in this manner, it is safe to say that other consumer devices could also leverage cloud-based assistance in their future. Alternatively, 'on device' solutions can offer users data security which delivers greater control over the data itself and where it might transit and finally reside.

The continued advancements in e-learning likely contributed to **56% of respondents saying that the education industry will increase their use and application of voice technology in the next 3-5 years.** e-learning was already a large use case for voice technology before the pandemic, but it was further accelerated due to the increased use of e-learning platforms and practices. Voice technology will continue to optimize these platforms by enhancing the accessibility of education through captioning in real time and transcribing lessons. With such a powerful potential to reach more individuals through learning tools, it's no wonder respondents see the huge potential for speech technologies in the education industry and expect to see these develop into the future.

> **"Everything starts from a conversation."**
>
> **Satvinder Nicky kaur Nagi**, Try Life

VOICE APPLICATIONS THAT WILL HAVE THE LARGEST COMMERCIAL IMPACT IN 2021

FIG 12. RESPONDENTS THINK THAT THE VOICE USE CASES THAT WILL HAVE THE LARGEST COMMERCIAL IMPACT WILL BE:



1.	44%	Web Conferencing Transcription
2.	37%	Customer Experience & Analytics
3.	27%	Subtitling & Closed Captioning
4.	27%	Education, Academic, Research Transcription
5.	22%	Medical Transcription
6.	20%	Consumer Electronics Command & Control
7.	20%	Automotive Command and Control
8.	17%	Chat App Messaging Transcription
9.	15%	Comms Monitoring
10.	14%	Media Monitoring
11.	12%	Voicemail Transcription
12.	10%	Legal Transcription
13.	10%	eDiscovery
14.	7%	Compliance

WEB CONFERENCING TRANSCRIPTION

44% of respondents think that voice technology will have the largest commercial impact on web conferencing transcription in 2021. Web conferencing has become a key part of our lives over the last year and this isn't expected to change. The pandemic has meant individuals have had to depend on these technologies not only in work environments but to stay connected to friends and family. Web conferencing tools have become well known and easy to use. With the most basic functionality often free, they are a staple communication channel.

Voice technology can further enhance web conferencing through improving accessibility with real-time captioning or deriving insights from calls by providing post-call transcripts. Web conferencing organizations have benefited greatly from the pandemic, placing huge demand on their services. Revenues have soared and with that has come increased budget to drive innovation. This popularity has also increased competition in the market with startups trying their hand at taking a chunk of market share. For this reason, web conferencing organizations wishing to remain leaders must invest in value-adding features like highly accurate speech-to-text to provide even more value to users.

CUSTOMER EXPERIENCE

37% of respondents said that voice technology will have the largest commercial impact on customer experience. Customer experience is a vital metric for many organizations for almost every application of product and technology. Simply put, if a customer has a bad experience with a product or service, they are unlikely to buy it again. Social media has made customer experience even more vital. It's never been easier to publicly share a bad – or good – experience and influence others, even if they have never interacted with the product or service directly.

Voice technology enables frictionless customer experiences. It helps humans interact with machines in an increasingly natural way – using their voice. The removal of physical user interfaces like keyboards makes it easier than ever for customers to make requests on the go. Voice technology is also helping brands to get a better understanding of their customers. By capturing all customer conversations and converting calls into text, brands can combine this data with text-based data obtained through emails, SMS, chat bots etc. From there, companies can understand root causes of customer issues and ensure that they don't happen in the future. This data can also be used to empower agents and ensure they are equipped to deal with customer issues and achieve first contact resolutions.

SUBTITLING & CLOSED CAPTIONING

27% of respondents believe that voice technology will have the largest commercial impact on subtitling & closed-captioning. It is expected that 2021 will see a continued surge in the creation of content as content providers and brands fight for audience eyeballs. Content is king and with uncertainty around lockdowns into 2021, brands are fighting to lock in existing customers and add new ones to their services through new and exclusive content.

Personalization is key to customer retention for streaming services and the ability to encourage the purchase of other services bundled into their OTT content package. In 2021, content curation will be more important than ever and the ability to generate quality metadata to extract key information about the media content will aid in offering a personalized service.

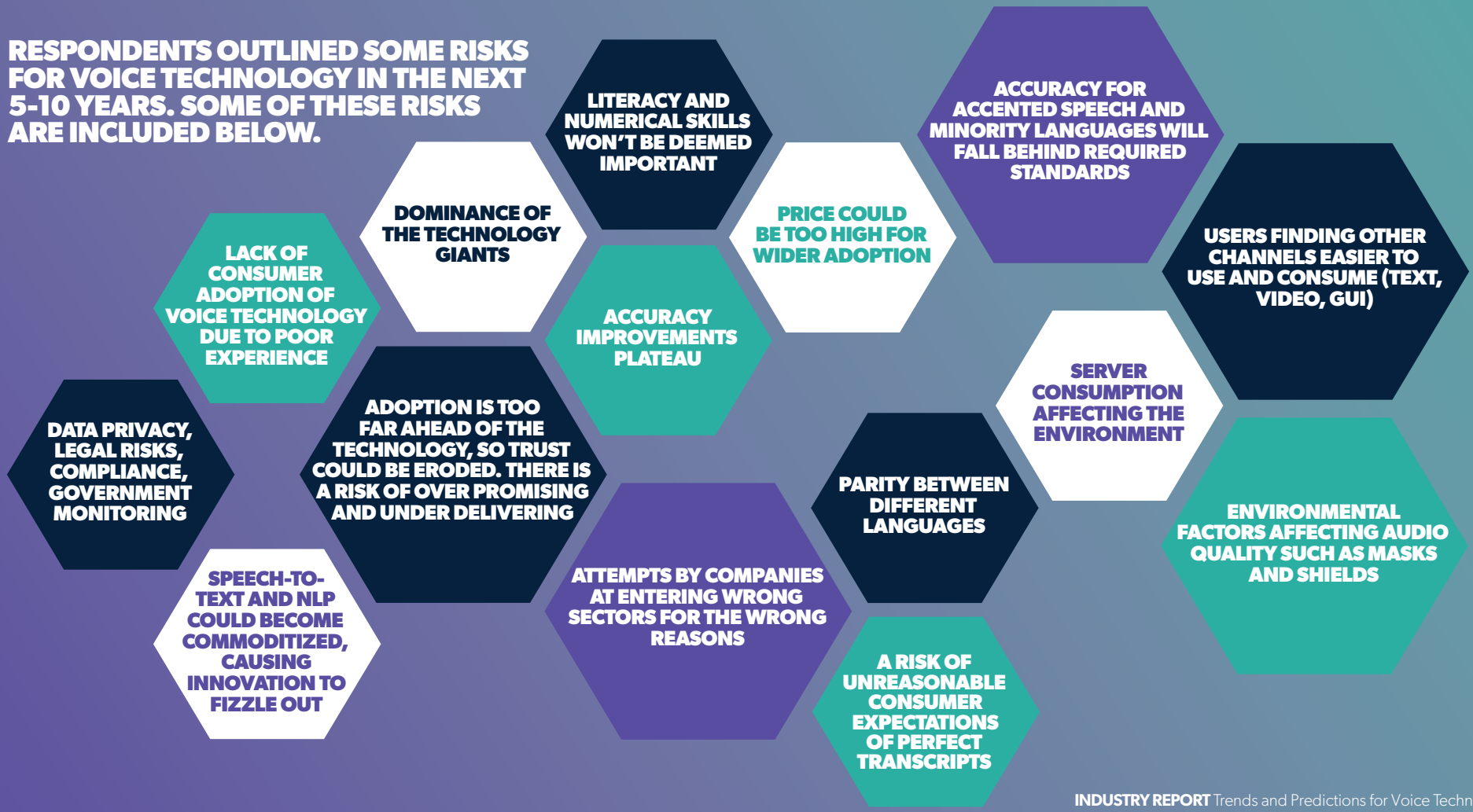
Brands and advertisers will continue to benefit from speech recognition technology to enable the management of their content catalogs. With fewer face-to-face interactions, brands must continue to drive the use of online video platforms, and voice technology is a core component of managing digital and media assets effectively.

Additionally, using vice technology to generate accurate subtitles and closed captions increases the accessibility of video assets. Whether deaf or hard of hearing, or situationally disadvantaged, captions drive video views and engagement.

Captioning is now demanded by legislation and will continue to be a challenge for organizations to deliver as content creation increases as expected. In 2021, organizations will need to improve their captioning capabilities and will adopt next generation tools like speech recognition technology. Additional features such as punctuation, speaker diarization and a diverse language offering will be key to help media teams keep up with the legislation brought about by the Federal Communications Commission (FCC).

RISKS FOR SPEECH TECHNOLOGY IN THE NEXT 5-10 YEARS

RESPONDENTS OUTLINED SOME RISKS FOR VOICE TECHNOLOGY IN THE NEXT 5-10 YEARS. SOME OF THESE RISKS ARE INCLUDED BELOW.



FUTURE CONCERNS AROUND DATA PRIVACY

95% OF RESPONDENTS SAID THAT DATA PRIVACY IS LIKELY TO BE A CONCERN IN THE FUTURE.

YES: 95%
NO: 3%
UNSURE: 2%

95% of respondents said that data privacy will be a concern in the future. People are concerned over where their data is stored and how. Data collection is a growingly important topic. Providing different deployment options for ASR – including on-premises – seems to be key to consumers being at ease with how their data is handled.

Consumers want to be able to access or maintain their data themselves and are looking to have complete transparency in the process. These concerns are an increasing trend given the track records of the tech giants such as Google and Amazon who freely use consumer's data. This makes users wary of the usage of personal data and how it could impact them. Consumers are also becoming savvier and want assurance and clarity to questions about how their data is being collected, stored, owned and utilized. Those conditions are influencing how users choose ASR providers.

"Historically EU-based financial services have been reluctant to engage with ASR due to onerous data localization and restrictions, but with on-premise solutions that comply with regulations likes GDPR, more are experimenting with speech technology solutions. I expect this trend to accelerate as contact centers and customer service use cases increasingly become the virtual front door of an organization."

Michael Tansini, Product Owner, Speechmatics

OVERCOMING THE CHALLENGES OF DATA SECURITY WHEN IT COMES TO VOICE TECHNOLOGY

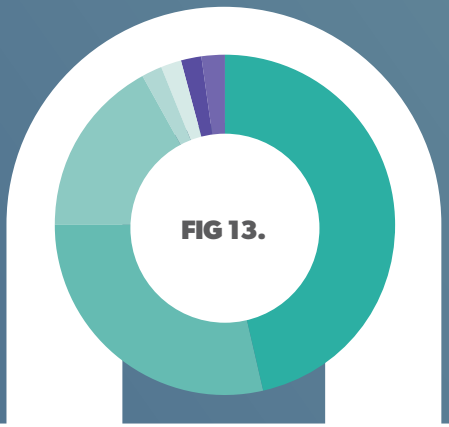


FIG 13.

1. **47%** On-premises deployment
2. **29%** Cloud deployment
3. **17%** Solutions able to operate entirely offline
4. **2%** Trust
5. **2%** Binding original recording with transcription securely
6. **2%** Anonymizing data and end users
7. **2%** Legal regulation in the cloud

Data security continues to be a concern across all industries. In 2019, concerns around voice data were prominent in the news with global brands such as Amazon and Google confirming they are using their home devices to "listen" to conversations for the development and improvement of their devices.

In 2020, these articles swiftly prompted a reaction from the tech giants to communicate the importance of data privacy. With such prominent voice data breaches, businesses are planning to overcome these challenges and ensuring data privacy is front of mind when integrating voice technology into their workflows.

We've established that 95% of respondents believe that data privacy will continue to be a concern in the future, but there will be ways to overcome data security issues. Professionals say that the way around it include:

ON-PREMISES DEPLOYMENT

47% of respondents said that on-premises deployment is a great way to overcome issues with data privacy now and in the future. This is an increase on 39% last year. On-premises deployment options enable users to keep their data secure within their own environments with no need for data to go into the cloud. On-premises deployments for voice technology are often done using virtual appliances or containers so they can be deployed effortlessly into existing technology stacks.

Industries such as banking, financial services and insurance face compliance and regulatory challenges where customer data and voice data cannot leave the premises to 3rd party providers. In these cases, on-premises deployments are the best solution to preventing data breaches and avoiding risks associated with private cloud deployments.

DARK SITE ENVIRONMENTS

Dark site deployment options enable customers to keep their data secure within their own data center environments. **17% of respondents said solutions that can operate entirely offline will alleviate their concerns around data security.** Typically, when deploying an on-premises solution for voice technology, businesses are required to connect to the public internet for licensing. Offline licensing is supported in dark site deployments meaning all work is completed within a business' private environment.

Offline licensing enables customers to license and operate the ASR solution without being connected to the public internet. This deployment delivers a more robust solution for compliance and data privacy needs.

Dark site environments are a great solution for many businesses such as governments that need an added level of security and privacy when it comes to voice data.

CLOUD DEPLOYMENT

29% of respondents said cloud deployments satisfy their business need for data security. Private cloud deployments are secure enough to keep data safe for lots of applications. If cloud deployment security is good enough for the business and use case needs, cloud deployment is often the preferred option due to low operational cost and complexities.

"I expect for large cloud providers to start offering on-premises deployments in the future as data privacy continues to be a concern for consumers."

Alex Fleming, Product Marketing Manager, Speechmatics

"The ability to do cost effective offline and on-device ASR is going to be critical to growth of many industries and use cases. It's a roadblock for everything from feature films to board meetings to healthcare."

Jim Tierney, Founder, Digital Anarchy

COMPETITIVE LANDSCAPE

The voice market has never been so competitive, with providers in the market as diverse as the applications requiring voice technology. Prominent tech giants such as Google, Amazon and Microsoft continue to extend their solutions in long-form speech recognition through the addition of new languages and enhancements to reduce word error rates. The continued adoption of devices like Google Home and Amazon Alexa have brought new audiences to their technologies now hungry to see the same level of command and control capabilities in other hardware such as automotive.

With individuals confined to their home as a result of the worldwide pandemic, the use of these products and solutions have seen even greater use than before. Virtual assistants have been used for information gathering around key topics as well as home automation use cases.

2019 saw Amazon making an effort to penetrate the medical market with voice technology which will likely see increased investment as 2021 unfolds. With immense pressure on this market through dealing with the pandemic, the

medical industry will look to advanced, innovative technologies that improve efficiencies and drive-down costs. With organizations such as Amazon focusing their efforts on the medical industry, there is a real opportunity to help transform the industry for the better through voice technology services.

As we predicted last year, Google also announced a product to 'bridge the gap' to an on-premises solution offering. This means voice services can be used differently to the traditional global cloud method of operation. The solution requires deep technical understanding to implement and operate, and so Google has targeted it towards advanced customers that are committed to a cloud and on-premises plan and spend.

Opportunities for voice technology as a result of the pandemic has also resulted in more players in the voice space. As the pandemic provided more opportunities for voice applications, more organizations emerged to offer these capabilities. It's expected that some of these solutions might take hold while others will likely either fall away or be consumed by bigger players in their markets. ASR remains a highly competitive market with still only a small number of players able to deliver enterprise-grade solutions that offer scale, accuracy and language capabilities. With levels of WER being similar across more providers for English, the ability to leverage machine learning to deliver increasingly sophisticated features to not only uplift transcription, but understand and extract extra key information will become essential.

VOICE TECHNOLOGY CONSIDERATION

WILL VOICE BE CONSIDERED IN YOUR 5-YEAR STRATEGY?

YES: 65%
NO: 35%

In the last year, more companies have adopted voice as part of their long-term strategies. **65% of respondents will be considering voice in their 5-year strategy.** Voice technology is reaching maturity and is becoming more widespread and accessible. Companies are realizing the value that voice technology brings to their businesses, not only to improve efficiencies and streamline workflows and processes but to improve customer experiences and ultimately increase revenues. As customer expectations increase, companies are encouraged to improve their products and services, and voice offers unique and innovative ways of doing this.

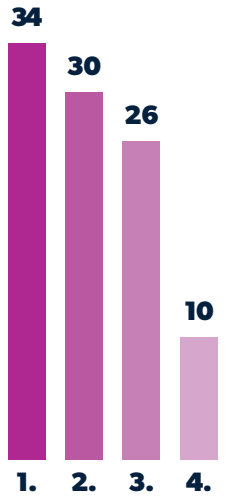
Emerging technologies such as natural language processing (NLP) and voice analytics tools are providing value to businesses across many industries. However, these tools can only be used with text as an input. This means, without the action of transcribing video, audio and voice into text, all of this data is left unusable. By converting these files into text, companies can offer a single source of truth for all of their data. With text and voice data combined and triaged, businesses can make use of NLP and analytics tools to obtain a 360-degree understanding of business processes and customer interactions, providing value where it has never been seen before.

"2020 showed us that more and more companies are realizing that voice technology is good enough to solve real-world problems and support significant improvements to workflows and customer experiences. The markets are waking up to the fact that voice technology is an enabler for them."

Alex Fleming, Product Marketing Manager, Speechmatics

GLOBAL REGIONS THAT WILL HAVE THE LARGEST GROWTH IN THE ADOPTION OF SPEECH RECOGNITION TECHNOLOGY

FIG 14.
(%)



1. **34%** Asia Pacific
2. **30%** North, Central and South America
3. **26%** Europe
4. **10%** The Middle East and Africa

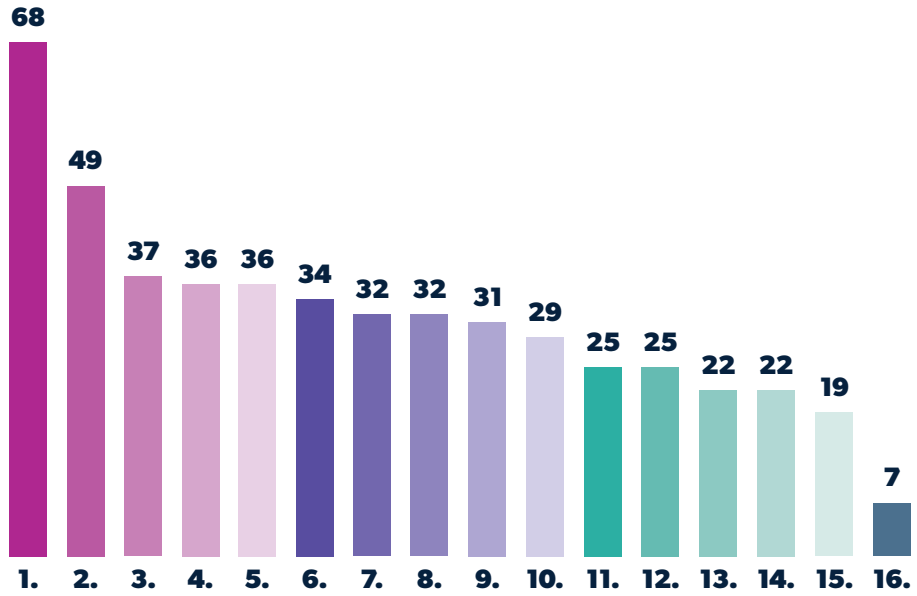
34% of respondents indicated that the Asia Pacific will have the largest growth and need globally for adopting speech recognition technology. This is largely due to the growth of the economy and population which will impact business and consumer trends. Following on from there, North, Central and South America, as well as Europe are also predicted to adopt voice technology at a rapid rate. Again, this is largely due to economic, social and technological factors, with the value of voice being realized. It is also fair to suggest that North, Central and South America and Europe are further ahead in voice technology adoption that Asia Pacific. Conversely, the Middle East and Africa are unlikely to adopt speech recognition technology rapidly as there is not yet a wide need for it in many instances.

LANGUAGE SUPPORT IN THE NEXT THREE YEARS

Over the next three years, respondents expect ASR language coverage to include the languages listed. As well as specific languages, respondents also acknowledge the need for any-accent language packs, for example, in the case of Spanish accents. Spanish is spoken in many countries across the world with varying accents and dialects, so an any-accent Spanish language model will become crucial to serving that market over the next three years. With so many languages having such a diverse range of accents within them, ASR will need to incorporate systems that are trained to be agnostic when it comes to deciphering and transcribing those languages to better serve users and increase accuracy in the longer term.

FEATURE DEVELOPMENT IN THE NEXT THREE YEARS

FIG 15. RESPONDENTS SAID THAT VOICE TECHNOLOGY FEATURE DEVELOPMENT WILL BE CRUCIAL OVER THE NEXT THREE YEARS, THESE FEATURES INCLUDE:



1. **68%** Improved word error rate (WER) accuracy
2. **49%** Increased speaker diarization accuracy
3. **37%** Language identification
4. **36%** Customer-specific language models trained on customer text data (language model adaptation)
5. **36%** Translation
6. **34%** Real-time transcription from the cloud
7. **32%** Short utterance accuracy
8. **32%** Increased number recognition accuracy
9. **31%** Non-speech detection (detect sounds, noises, music, disfluencies, hesitation, silence)
10. **29%** Customer-specific language models trained on customer acoustic data (acoustic model adaptation)
11. **25%** Noise reduction
12. **25%** Audio file quality assessment
13. **22%** Increased speed/latency
14. **22%** Word alternatives available in the output
15. **19%** More languages available
16. **7%** Redaction

IMPROVED WORD ERROR RATE ACCURACY

68% of respondents said that they would like to see better accuracy over the next three years. WER will continue to improve by throwing more data at the problem, however, this approach will likely see diminishing returns. The leading providers in the ASR space continue to deliver WER accuracy of around 95% for English. Using data to increase accuracy even further will require a huge amount of data and increasing levels of processing power for single percent increases. For many providers, this will not be worth it, or they simply will not have the available hardware to retrain with such huge data sets. With English reaching a point of accuracy which is hard to surpass, ASR providers need to look at the accuracy of other languages and ensure they are fit for purpose for global businesses. Providers can also look to shift focus to delivering supporting features that enhance the quality of output provided to its users.

Features like entity tagging within the audio, identification of the languages spoken and better diarization will all count towards the delivery of a more accurate representation of the audio as part of the files transcribed. Providers also need to ensure that the levels of accuracy that they preach are applicable in real-world use cases. For example, the ability to deliver quality transcription output in noisy environments or with audio recorded on low-quality devices.

WER improvements will likely be incremental compared to the past few years. However, with issues around bias continuing into 2021 and beyond it's likely that improvements can be made across all languages, particularly when it comes to accents and dialects.

SPEAKER DIARIZATION

49% of respondents said that they would like to see better speaker diarization accuracy over the next three years. Speaker diarization is used to understand which speaker was talking in single-channel media files. It does this by detecting unique speakers and assigning speaker labels to the corresponding portions of text within the transcript. It's one thing to know what was said, but the means to split the transcript by speaker adds additional value to a range of users.

Speaker diarization is one of the most challenging elements of speech recognition. While speech and other audio characteristics are easy for the human brain to detect, this poses a challenge for automated systems due to the fluctuations in a single speaker's voice depending on their mood, hesitations, word emphasis, noise etc. While speaker diarization exists today, it is still a key challenge that speech providers have not yet mastered. 2021 will likely see increased effort to improve speaker diarization to uplift use cases that benefit from being able to match a speaker with the words spoken.

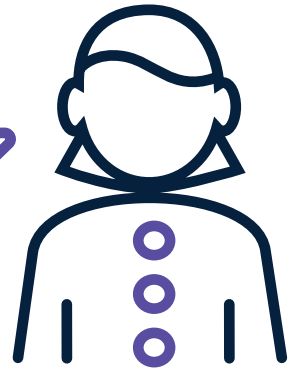
LANGUAGE IDENTIFICATION

37% of respondents said they would like to see better language identification over the next three years. Being able to generate more information about a file and trigger additional actions off the back of this information is becoming more popular. Detecting the language of the speakers within a video or audio file automates the manual task of selecting the correct language pack to use to transcribe it.

ASR providers are continuing to diversify and are perfectly positioned to deliver additional capabilities due to the access to data that they have. By automating the language identification element of the transcription process, businesses can save time and human resource cost as well as unlock new information that would previously have been lost. For example, in stock trading where compliance and monitoring are vital, the means to understand that a call might contain multiple languages might identify it for additional investigation, not just help select the right language pack to use to transcribe it.

"WER will be similar across providers and so new elements will be required to differentiate the best solution for each person, company and use case. There will be demand for understanding what ASR can deliver and the power of it rather than just a transcript."

Alex Flemming, Product Marketing Manager, Speechmatics



CUSTOMER-SPECIFIC LANGUAGE MODELS TRAINED ON CUSTOMER TEXT DATA (LANGUAGE MODEL ADAPTATION)

36% of respondents said that they would like to see customer-specific language models trained on customer text data (language model adaptation) over the next three years. Customization of language models have existed for many years. Users can import their word lists into ASR engines to recognize brand and personnel names, acronyms and other words that are likely to be absent from the training data used to create language packs. While this approach has the potential to significantly help in the recognition of certain words, it lacks the finesse of context.

The ability to tune models using the user's own data has the potential to deliver the extra and elusive 5% in WER accuracy that standard packs might not. While custom language models might deliver the accuracy required by some users, they come with some challenges. Providers like Microsoft require users to hand over their data in return for tailored language offerings, however, this breaches data confidentiality to a point that is often unacceptable to many ASR users.

Users looking to take the next step in their ASR journey will be required to work more closely with their providers to collaborate and transcend a customer/provider relationship. It will require more closely partnering through sharing data and improving incrementally until joint goals are achieved.

SHORT UTTERANCE ACCURACY

32% of respondents said they would like to see better short utterance accuracy over the next three years. With increased focus on virtual assistants and as they continue to be applied to more edge-based platforms like phones, cars, and other devices, it's no surprise that respondents expect accuracy improvements in short utterance accuracy in the future. Traditional static systems based in homes or work places have their own issues when it comes to understanding and interpreting what you said due to noisy environments, accents and unclear commands.

The global adoption of virtual assistants has encouraged providers to continue to add high-quality ASR in even more languages, for more accents and dialects than ever before. Consumers expect their virtual assistants to understand them irrespective of their accent, dialect or language. In fact, 19% of respondents called language coverage out as something they expect providers to improve on in the next three years.

SPOKEN LANGUAGE TRANSLATION

36% of respondents said they would like to see translation over the next three years. Innovation within voice technology means that industry use cases will continue to evolve with an expectation that speech recognition accuracy will improve, and features and intelligence will also grow around it.

Organizations with aspirations to deliver a global service must unify communications and messaging across all their work environments and employees irrespective of location. Translation might provide an answer to this; however, it presents challenges that still require work to solve. Audio can be transcribed in one language, translated word for word, and then fed into a text-to-speech engine. The output, however, will never reflect a natural language. To achieve results in this application, additional understanding and experimentation will be required with specialist providers to dedicate effort to enabling the delivery of a transcribed, translated and machine spoken output that is near to indiscernible from a natural speaker.

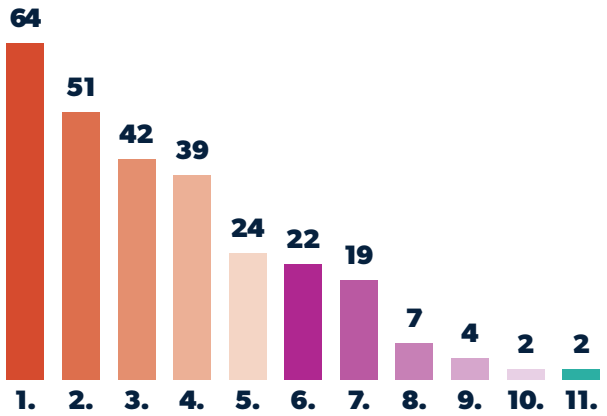
"Speech will only realize its potential when ASR can parse the semantic detail latent in even short utterances, whatever the intent of the utterance. So, it's very close to being a hard-AI problem."

Iain Mackay, Director, x-mr

PERCEPTION OF ACCURACY

The perception of accuracy for voice technology is very specific to the use case and business. There are several measures that respondents said are important when it comes to accuracy, which you can see in Figure 16.

FIG 16.
(%)



1.	64%	Word error rate
2.	51%	Intent recognition
3.	42%	Speaker change indicated
4.	39%	Punctuation
5.	24%	Multiple languages
6.	22%	Quick turnaround time for transcription
7.	19%	Unified number recognition format
8.	7%	Character error rate
9.	4%	Robustness to background noise
10.	2%	Speaker identification
11.	2%	Voice tone analytics

The perception of accuracy for voice technology is often interpreted as the word error rate (WER). **64% of respondents also had the same perception**. WER is the industry standard for measuring the accuracy of word output in a transcript and is a great benchmark. However, many businesses and use cases use other metrics to measure the accuracy of the transcription output.

42% of respondents said speaker change being indicated in the transcription is an indicator of accuracy. Understanding the conversation flow provides an added level of accuracy to the transcript other than just how accurate the word output is. Industries such as contact centers and media, entertainment and broadcast find this feature incredibly valuable in real-time scenarios as well as post-processing.

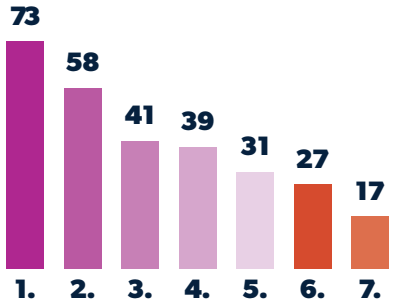
51% of respondents also said intent recognition is an important feature in an accurate transcript output which is up from 41% in 2020. Intent recognition provides the context of what has been said and can often change the output of a sentence to ultimately increase the accuracy, but this could directly affect the WER. This increase in importance reflects how the industry is moving towards understanding the output rather than just the accuracy of the words.

"In the age of Robotic Process Automation, failure to capture the different accents and domain specific lingo means manual revision and as a major hit to the added value of the solution."

Diego Montoliu, Director, Delta-ai

THE FUTURE OF SPEECH RECOGNITION

FIG 17. RESPONDENTS SAID THAT THERE ARE LIKELY TO BE CHANGES IN THE SPEECH RECOGNITION MARKET. IMPORTANTLY, THE CHANGES WILL BRING IMPROVEMENTS TO THE TECHNOLOGY THAT WE SEE TODAY. SOME OF THOSE KEY IMPROVEMENTS AND CHANGES ARE OUTLINED BELOW.



1. **73%** Improved word error rate (WER) (95%+)
2. **58%** More robustness when it comes to noisy environments
3. **41%** More focus on any-accent language models
4. **39%** Less data required to improve models
5. **31%** More on-device deployments through reduced model size
6. **27%** Enterprise-grade ASR is no longer just speech-to text
7. **17%** Large companies will offer more deployment options

73% of respondents said the future of speech recognition is improved word error rate (WER). Looking back to the perceptions of accuracy, low word error rates are regarded to be the main definition of accuracy. As machine learning algorithms continue to evolve, it is likely that WER accuracy will reach 95%+ especially for commonly used languages like English. However, there is still significant work to do to achieve this due to the range of accents and dialects present in all languages and to deliver the same level of accuracy across them all. ASR providers and users need to be objective around how they conduct testing to ascertain these WER scores and understand what this means for them and their customers. In some cases, WER might also fall to the wayside when compared with other KPIs such as speed.

58% of respondents said the future of speech recognition should see more robustness when it comes to noisy environments. It's no surprise that the ability to deal with noisy environment is a key factor in the future of speech recognition. Noise is a major factor that can impact accuracy. It was also highlighted as a risk that we could experience a downfall in audio quality due to the pandemic requiring the population to wear masks and other protective wear. Being unable to detect the words of a speaker due to noise has a direct impact on the outcome of the transcript, so the ability to reduce interference or deliver high-quality recognition even in challenging environments remains a top priority.

Contemporary ASR solutions are incredibly effective even in noisy environments or when media is recorded on low-quality devices. ASR providers are likely to continue to improve the diversity of their training models to include challenging audio profiles to deliver greater robustness when dealing with noise. Aggressive benchmark testing that identifies key areas for improvement will also help to identify key areas that can be fortified to ensure quality of ASR even in challenging audio environments.

Most ASR providers have multiple accent packs for their languages. Due to this, the expectation by **41% of respondents is that the future will see a greater focus on accent-independent language packs**. It is expected that providers who support accent-specific packs for the same languages will continue to support this strategy even as accents continue to diversify and broaden. Brands might have to accept aa slowdown in updates to specific language packs from providers with an increasingly global demand. Additionally, providers will need to consider the impact of balancing the cost of deploying and operating an increasing number of language packs for single languages and how they can make the right decision on which to use for each use case and for each file.

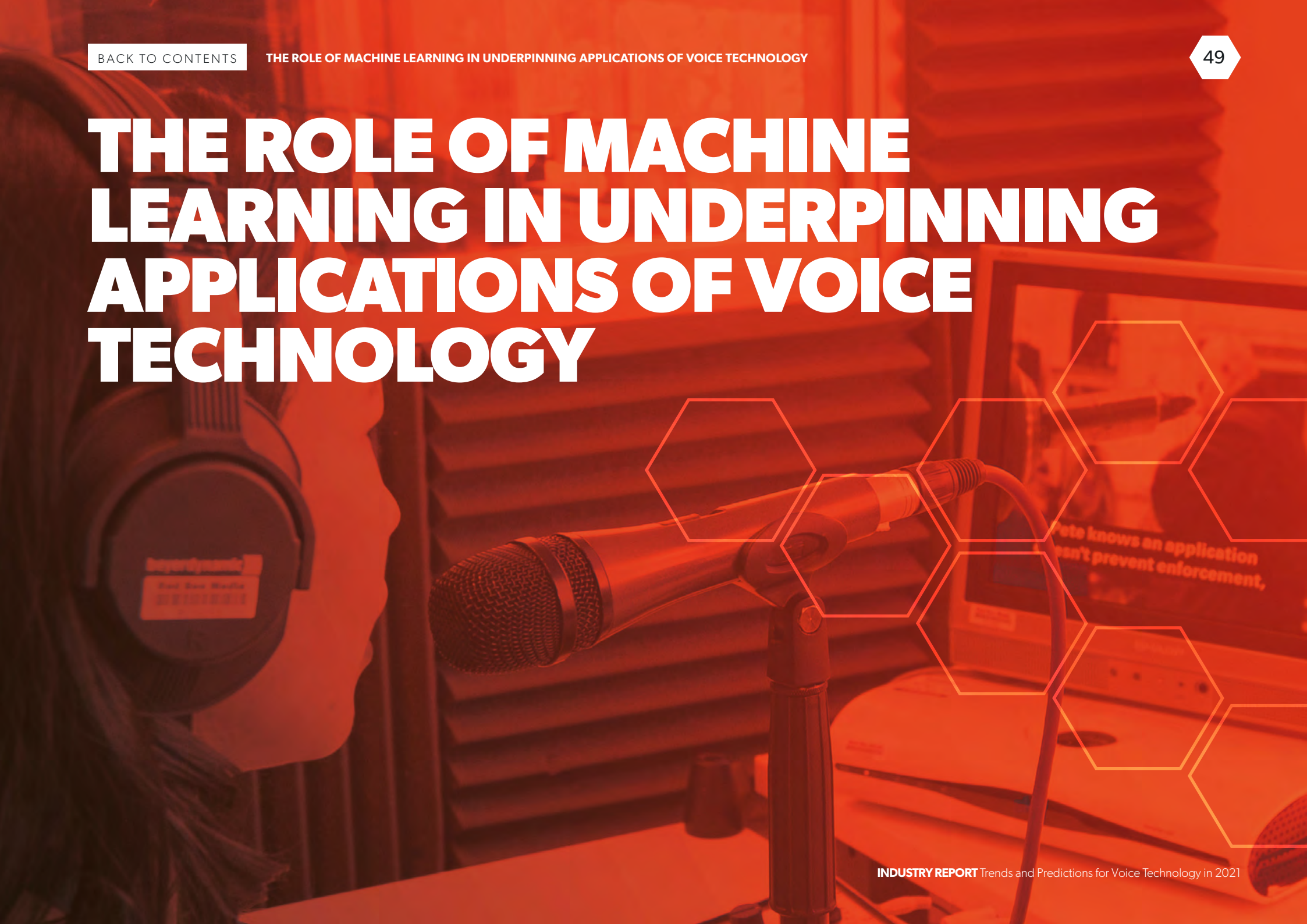
Regarding languages used by smaller populations of people such as Irish, Icelandic and Estonian, it is likely that these languages will remain without a dedicated ASR model for some time. For this reason, users in the countries or areas that these languages are spoken will likely be forced to use another language to interact with devices and applications that are voice-enabled should they wish to.

27% of respondents believe enterprise-grade ASR is no longer just speech-to-text functionality. Where previously, ASR and speech-to-text meant the same thing, as time has gone on and new features and capabilities have been added and become the norm, it now delivers far more than just transcription. In the future, it is expected that transcription (the core function of speech-to-text) will become just a single component of the wider ASR offering. The addition of punctuation, diarization and language identification, are all additional elements that optimize the speech-to-text component of the overall solution, creating an output that is more in line with what might be expected from a human.

"Speech recognition has to move away from simple data analysis or pattern matching, and start to include a level of understanding, at least of dialogue structure and intentions."

Alex Monaghan, Director Presales EMEA, Eckoh

THE ROLE OF MACHINE LEARNING IN UNDERPINNING APPLICATIONS OF VOICE TECHNOLOGY



NEW MACHINE LEARNING METHODS THAT ARE BEING UTILIZED

Machine learning underpins the progress of voice technology. The approach of applying recurrent neural networks to speech recognition was demonstrated in the 1980s, where it outperformed traditional methods. The rise in computing power, graphics processing and cloud computing made the huge potential of this approach a reality.

Advances in machine learning, new techniques and innovation are all contributing towards continuous improvements in speech recognition technology, not only through providing better accuracy but also in feature development and language capabilities.

Already, voice providers are using machine learning innovation to provide value to consumers through better language capabilities. For Speechmatics, this means doing away with specific language models for different accents i.e. UK, US, AUS, and providing a single English language pack. As predicted in our 2020 report this method has been extended to Spanish to deliver some of the highest accuracy in the market across a range of global Spanish accents and dialects from a single language pack.

"Self-supervised learning will be able to extract increasing amounts of value from unlabeled data and drive improved generalization. Whereas semi-supervised learning is proving itself as an effective way to use unlabeled data to adapt to particular domains – iterative pseudo-labeling schemes have shown great promise recently and show good scaling properties."

Will Williams, Machine Learning Engineer, Speechmatics

UNSUPERVISED LEARNING

There has been a lot of progress in the machine learning field with natural language processing (NLP) due to a family of algorithms called unsupervised learning. Companies are being built based on the breakthroughs in unsupervised learning in text and language, but those algorithms can also be applied to speech. So, how can companies label elements in speech without using labeled data? For this to work, unsupervised learning algorithms need to be leveraged to see the same breakthroughs we've seen in NLP but to bring those through to speech recognition. This is a trend that is starting now and we're going to see a lot more of it in 2021.

To effectively train machine learning models for use in speech recognition, thousands of hours of labeled data is required. The issue with labeled data is that it is very expensive, hard to get hold of (especially in all the required domains and languages) and is often mislabeled. Unsupervised learning mitigates against these problems and enables machines to learn much more like humans would, with limited labeled data.

END-TO-END STYLE APPROACHES

End-to-end style approaches to ASR will continue to be researched to remove the need for expert knowledge in this field. An end-to-end approach is still a way off and is likely to have an impact in the next 5-10 years. In the meantime, network distillation, quantization and other technical approaches will be used to reduce model size to enable more deployment options, especially on-devices.

SUMMARY

2020 was a year like no other and it's likely the hangover from the pandemic will be felt for many years to come. One thing that 2020 taught us was how hard it is to predict the future and a global pandemic was certainly not one of our 2020 predictions.

In recent years, voice technology has seen an upward trajectory in both popularity and adoption. Voice technologies are no longer the thing of science fiction, they are a valuable asset that see no sign of slowing down in the foreseeable future. Early adopters of voice technologies were able to leverage these immediately when the pandemic hit, augmenting teams and processes in an attempt to combat the demand. Those who had waited, in some cases, were forced to adopt technologies such as voice in reaction to the challenges they faced in their new remote workforce. Telecommunications was highlighted as an industry that has experienced one of the biggest positive impacts as a result of COVID-19 through the mass adoption of solutions such as Zoom, Webex and Microsoft Teams. It is likely that without these already mature – but at the same time – lesser-known solutions, that working through the pandemic would have been significantly more challenging and would have had an even greater impact on society and the economy as a whole.

From consumer devices in homes like the Amazon Echo and Google Home to the deployment of voice-based solutions within business and enterprise environments, voice has and will continue to be additive to our daily lives.

50% of professionals said that their company currently has a voice strategy. Further to this, 70% said that voice is likely to be considered in their business' 5-year strategy. This emphasizes that voice will continue to be a core technology to drive the advancements in businesses across many industries.

Over the last 5 years, the application of speech technology within voice bots and virtual assistants has continued to grow and our research shows that there is no sign of this stopping. Virtual assistants remain the most visible application of machine learning, artificial intelligence and speech recognition technology on the market. With that said, terms like AI are often used very broadly and on occasions misused to describe a technology landscape to make it sound impressive. This presents a risk in the enterprise space to stifle adoption of genuinely groundbreaking technology due to a poor experience with solutions that claim lots but delivers very little.

The use cases around voice are also becoming increasingly complex dependent on industry, market, use case and consumer base. While cloud providers might provide the easier route to voice-enabled products or services, considerations should be made to understand if that approach is the right decision to meet consumer and business needs in an evolving market that cannot be predicted.