

# Acceptance of Speech-to-Text Technology: Exploring Language Proficiency and Psychological Safety in Global Virtual Teams

Carolin Fleischmann  
Ansbach University of Applied Sciences  
[carolin.fleischmann@hs-ansbach.de](mailto:carolin.fleischmann@hs-ansbach.de)

Jolanta Aritz  
University of Southern California  
[aritz@marshall.usc.edu](mailto:aritz@marshall.usc.edu)

Peter Cardon  
University of Southern California  
[cardon@marshall.usc.edu](mailto:cardon@marshall.usc.edu)

## Abstract

*Natural Language Processing, such as speech-to-text technology, is increasingly implemented in collaboration software that is used by global virtual teams (GVT). GVT collaboration has become ubiquitous and has additionally accelerated during the COVID-19 pandemic. The main issues of global virtual teams are technology difficulties, language and time zone differences, and lower levels of psychological safety. Advances in collaboration technology aim at improving collaboration for GVT. But we know little about the acceptance of these technologies. Therefore, the objective of this study is to explore how Millennial and Gen Z members of GVT accept speech-to-text technology; namely, automated captions in virtual conferences and automated meetings transcripts. Particularly, we are comparing antecedents of acceptance across levels of language proficiency and psychological safety. We surveyed 530 users of speech-to-text technology in GVT both before and after they used the technology. The pre-survey was administered before the COVID-19 pandemic hit; when participants completed the post-survey all were under some degree of lockdown. Results suggest that use of the technology reduces anxiety and effort, but decreases performance expectation and hedonic motivation. Non-native speakers rate the technology more positively. The impact of psychological safety is limited to self-efficacy and anxiety.*

## 1. Introduction

Communication technologies increasingly implement natural language processing (NLP) capabilities, such as automated captions and transcripts for online meetings [1]. These speech-to-text technologies promise easier collaboration and support for completing tasks. Transcripts, for examples, can make meeting minutes obsolete, thus freeing up time that can be used for other tasks. Yet, the use of

emerging virtual collaboration technologies has received little research attention [2].

These potential benefits can only be leveraged when users accept the technology. Companies that commercialize these technologies as well as companies that use them in their online meetings need to know more about speech-to-text technology acceptance and use this knowledge to further develop the technology.

Whether users accept the technology will depend on how they respond to the technology itself as measured by the Unified Theory of Acceptance and Use of Technology (UTAUT) models [3] [4]. In addition to that, technology acceptance will depend on boundary conditions on the individual and interpersonal level.

Language proficiency, for example, is a major topic for global virtual teams (GVT) and has been found to impact preference for certain communication media. Lower proficiency speakers are less likely to prefer rich communication channels such as online conferences. They benefit from written communication [5] [6].

On an interpersonal level, we include psychological safety as a boundary condition because transcripts are often accompanied by the fear of private conversations being made public. Therefore, a safe environment may increase the likelihood of speech-to-text technology acceptance. So far, trust has been researched in the context of NLP (e.g. [7] [8] [9]), but related concepts, such as psychological safety, lack investigation.

Research on the acceptance of NLP technology remains scarce, particularly in collaborative settings [10]. Some studies have analyzed the use of voice assistants and other NLP technologies [11] [12] [13] [14]. A study of automated transcripts and captioning will complement these findings to create a more complete view on NLP technologies.

Our study on automated transcripts and captioning will help answer questions around their acceptance: How will members of GVT accept speech-to-text technology? Will automated transcripts and captions help GVT, and particularly non-native speakers of the team language, perform better? Will different levels of psychological safety influence

technology acceptance?

Thus, the objective of this study is to investigate speech-to-text technology acceptance across different levels of technology use, language proficiency of users and psychological safety.

To test our hypotheses we surveyed 530 team members of 90 global virtual teams. Participants completed two surveys: one before they used automated captions and transcripts in their virtual teamwork and another one after they had used the technology for seven weeks. The pre-survey was administered in February 2020 before the COVID-19 pandemic impacted participants' lives. When participants completed the post-survey in April 2020, all of them were under some degree of lockdown and worked remotely. A significant number of participants were located in hard-hit areas, such as Spain or New York.

The contribution of this study is two-fold: For research, we expand the body of knowledge on technology acceptance by including language proficiency and psychological safety as boundary conditions. We provide empirical evidence for the role of these boundary conditions for technology acceptance. In particular, we provide insights from a time when technology-mediated communication suddenly became the primary mode of communication due to stay at home orders.

For practice, we contribute nuanced insights on the current level of speech-to-text technology acceptance among Millennial and Gen Z users. Our research indicates that multilingual settings may be a promising context for speech-to-text technology commercialization. Leaders and managers of global virtual teams should consider adopting captioning and speech-to-text tools to facilitate better communication when non-native speakers are present. For all virtual teams, leaders and managers should consider the benefits of captioning and transcripts to aid in creating minutes and enhancing accessibility.

## **2. Theoretical background and hypothesis development**

### **2.1. Speech-to-text technology for global virtual team collaboration**

Speech-to-text technology is part of the family of Natural Language Processing technologies, where machine learning is used to detect, understand, analyze, and act on human language text or speech. Most often, these technologies help find information and reply to requests of users [15]. The technology works and is accepted by users if it provides "the right information at the right time" [16]. The most prevalent

examples of these technologies are probably Apple's Siri and Amazon's Alexa that use human spoken language as the user interface.

In speech recognition, which is the basis of speech-to-text technology, the machine captures spoken human language as raw audio data from the microphone. The machine learning based technology converts audio to text based on an algorithm that processes short sequences of audio to predict the most likely word. While some differences exist, the predictive model usually takes learned context and the actual audio cue into consideration [17].

In video conferencing tools, speech-to-text technology is used to provide automated captions in real time or meeting transcripts after the meeting has ended. Zoom, for example, has integrated the transcription software Otter.ai to auto-transcribe recorded meetings.

The quality of captions and transcripts is benefitting from the fast advancements in predictive capabilities, which expand the technology's use cases and acceptance levels among users. Captions and transcripts can facilitate understanding and improve accessibility in online conferences; transcripts have the additional benefit of providing a sharable and searchable log of the meeting conversation, which can be helpful for team members who missed the meeting (e.g. for internet connectivity or time zone issues) and non-native speakers who have trouble following all conversations in real time [6].

On the downside, captions may distract meeting participants; transcripts bear privacy concerns and still have some trouble distinguishing between speakers. Both are still lacking accuracy despite their increase in quality over the last few years [18].

In our study, the GVT used the auto-captioning capabilities of Skype and auto-transcripts by Otter.ai. Skype has been offering real-time auto-captioning since 2018, where users see captions in the Skype conference as they speak. Otter.ai is a Silicon Valley based start-up that launched their NLP software in 2018. Users receive a transcript of the video conference shortly after the end of the meeting. Otter.ai is licensing their technology to Zoom as well as selling directly to users. Since Otter.ai has integrated in Zoom and particularly during the shift towards work from home during the COVID-19 pandemic, meeting transcriptions have become increasingly prevalent.

### **2.2. Technology acceptance model for speech-to-text tools**

To make use of the benefits of speech-to-text technology in global virtual teams, team members

need to be willing to accept the technology. We assess the antecedents of speech-to-text technology acceptance based on the widely used and robust UTAUT models [3] [4]. Two versions of the Unified Theory of Acceptance and Use of Technology exist. We combine the benefits of both – the organizational context from the original UTAUT model and more differentiated variables from UTAUT 2 – to fit the needs of our study. For our particular case, technology acceptance is a function of performance expectancy, effort expectancy, hedonic motivation, self-efficacy, and anxiety. We excluded social influence, facilitating conditions, habit, and price value because the study setup held these variables constant.

We measured the UTAUT variables before and after teams used the technology. Performance expectancy measures how useful someone thinks a technology will be for accomplishing a goal. It has a strong influence on behavior intention [3] [4] [19]. For speech-to-text technology in virtual teams, individuals will assess whether the technology will facilitate understanding and overall communication in the team.

Effort expectancy emphasizes the process of using the technology rather than the outcome. It measures whether users believe that the speech-to-text technology will be easy to use and integrate seamlessly into the team's communication infrastructure. Team members will rate effort expectancy higher when they feel that using the technology causes additional work and disrupts team communication.

Hedonic Motivation represents the intrinsic motivation to use a technology – whether it is enjoyable to interact with the technology [11] [20]. Team members assess how much fun it is to use speech-to-text technology.

Self-efficacy measures how well a user can navigate technology challenges. Team members rate speech-to-text technology according to their individual ability to troubleshoot.

Finally, anxiety measures if a user is scared of intimidated by the prospect of using the technology [4]. Speech-to-text technology may create anxiety for certain team members, particularly for those with little experience with these kind of tools.

How individuals accept a technology depends on familiarity with the technology, where acceptance increases with familiarity. Empirical evidence exists that anxiety and apprehension about using technology decreased after GVT used the tools [21] [22]. While speech-to-text technologies are widely employed for personal use, virtual team collaboration tools are only starting to include the technology. Therefore, familiarity is limited, and we hypothesize:

*Hypothesis 1: Speech-to-text technology*

*acceptance in GVT will be higher after teams have used the technology.*

### **2.3. Language and psychological safety in global virtual teams**

GVT rely heavily on technology to communicate [23]. Therefore, advancements in communication and collaboration technology have the potential to change the way these teams work together. Whether GVT accept these new technologies, such as speech-to-text tools, depends on several factors.

Some of these factors are unique to virtual teams and rarely encountered in collocated teams. In global teams, individual team members often have different levels of language proficiency, which influences their preferences for certain communication technologies [6]. In addition, virtuality often leads to a lack of informal interpersonal interaction. Thus, GVT have more difficulty developing interpersonal relationships, including psychological safety [12].

With regards to language barriers, research has shown that non-native speakers are less likely to prefer spoken communication, i.e. video conferences, for decision making in their team. Written communication gives them the opportunity to revisit message content as much as needed (reprocessability) [5] [24]. However, written communication is not a suitable replacement for video conferencing. Those teams that hold more video conferences have an advantage in creating a feeling of inclusion and satisfaction with their teamwork [24]. Additionally, the use of rich communication media increases team performance in highly diverse teams [25].

Therefore, a combination of spoken and written communication may remedy the disadvantages of both communication media choices. While team members still engage in rich communication via video conference, they are better able to follow the conversation through the use of captions. The technology empowers non-native speakers [1] because reading a foreign language is often easier than understanding spoken language; native speakers have reported that Skype automated captions understand even strong accents that are difficult for native speakers to decode. Performance expectancy for speech-to-text technology is likely high.

For transcripts, non-native speakers benefit from being able to reprocess a meeting – or parts of it – that was difficult to follow. Transcripts offer all team members the chance to detect and follow up on misunderstandings and add additional context to meeting contents if necessary. While captions and

transcripts can be useful for all team members, non-native speakers will particularly benefit from their capabilities [6]; thus, they will have a high performance expectancy and therefore be more likely to accept the technology.

Thus, we hypothesize:

*Hypothesis 2: In GVT, speech-to-text technology acceptance is higher for team members who are non-native speakers of the team language.*

While benefits of speech-to-text technology clearly exist, it may create anxiety for some team members because they have privacy concerns if everything they say is recorded, transcribed and potentially shared with outsiders [26]. This may lead to restrictive information and knowledge sharing, inhibit learning, low performance expectancy and hedonic motivation, and ultimately lead to a rejection of speech-to-text technology for collaborative settings. Team members that perceive their team as a safe environment, where they don't need to fear negative consequences after they have openly shared their opinion in meetings, are less likely to perceive such threats.

Psychological safety is defined as a “shared belief held by members of a team that the team is safe for interpersonal risk taking” [27]. It includes an assessment of the team environment as to how others will react to seeking feedback, pointing out mistakes, or pitching ideas [27]. If the team environment is perceived as non-threatening and no negative consequences are expected when expressing oneself, a person will feel psychologically safe [28]. While the concept is related to cohesion and trust, it is different: In contrast to cohesion, psychological safety implies confidence to disagree rather than group think. Trust is other-oriented (“Can I trust *you*?”); whereas psychological safety is self-oriented (“How will others respond to *my* behavior?”) [27].

By fostering feedback and a culture of reflecting collaboratively on mistakes, psychological safety has been associated with learning and knowledge sharing [27].

Similar results have been found for virtual environments (e.g. [28], [29], [30], [31]), but studies on psychological safety in virtual teams that never meet in person remain scarce [32]. Gibson & Gibbs (2006) found that virtuality of collaboration has negative effects on innovation, but this negative effect can be mitigated by psychological safety [29]. Zhang et al. (2010) found that psychological safety increases the intention to continue sharing knowledge in virtual communities [28]. Kirkman et al. 2013 found that psychological safety increases team performance, particularly for teams with high national diversity.

National diversity is the norm in the GVT in the sample of this study [25].

In an environment that feels psychologically safe, team members may be more likely to accept meeting transcripts because mistakes and comments in meetings are not expected to be held against someone. Transcripts and captions are rather seen as positive because they allow for better feedback structure and provide opportunities for learning and knowledge exchange.

Therefore, we hypothesize:

*Hypothesis 3: Speech-to-text technology acceptance is higher for team members that feel psychologically safe in their GVT.*

With their functionalities, captioning and transcripts have the potential to combine the benefits and remedy the disadvantages of different existing communication technologies: Team members can build interpersonal relationships in rich communication channels, such as video conferences. At the same time, language barriers are decreased due to adding a written element to video conferences by captioning and transcribing them. When team members feel psychologically safe, they are more likely to rate anxiety with the technology low and see the performance and hedonic motivation of using speech-to-text technology.

### 3. Methodology

#### 3.1. Sample

The survey was deployed in global virtual teams that completed a consulting project for a large multinational organization in the technology, hospitality, or automotive industry. Teammates were dispersed around the globe, never met in person during their 7-week project, and did not know each other nor had worked together before the project. They used Skype with automated captions for video conferencing, Otter.ai for transcripts, and Slack for written communication and file sharing. Team produced a report of their analysis, findings, and recommendations for clients.

Project participants were Millennials and Gen Z that were enrolled in an undergraduate or MBA program at one of 16 participating institutions in seven countries on three continents. In total, participants in the study come from 38 different countries originally. 53% of participants considered themselves native or near-native speakers of English, which was every team's working language. Non-native speakers had at least a working level of English proficiency that

allowed them to participate in the team work.

A total of 530 individuals in 90 teams participated in the project. All teams had 5 or 6 team members and were composed to reflect similar levels of diversity. Each team had 40-60% US-based team members from different institutions across the country. The other team members came from institutions in Finland, France, Germany, India, and Lithuania.

Data was collected via quantitative surveys before the teams began working together, in February 2020; and after the project had ended in April 2020. Pre- and post-project survey responses were matched using person-specific identifiers. In addition to quantitative measures, participants were asked to share qualitative comments on speech-to-text technology in open-ended questions. These comments were analyzed using a NLP-based sentiment analysis tool. The response rate was high at 78.1% for the pre-project survey and 80.4% for the post-project survey.

### 3.2. Measures

To test our hypotheses, we measured (a) antecedents of technology acceptance pre-project, (b) antecedents of technology acceptance post-project, (c) language proficiency, and (d) psychological safety.

We measured antecedents of speech-to-text technology acceptance using measures from the UTAUT models [3] [4]. The study design held some of the UTAUT measures constant. Therefore, we focused on a subset of the variables; namely, performance expectancy, effort expectancy, hedonic motivation, self-efficacy, and anxiety. These constructs were measured at two points: before the teams began working together and using speech-to-text technology as well as after the project ended.

Performance expectancy was measured on a 4-item-scale from the original UTAUT model [3]. The original UTAUT model measure was chosen because it includes an item on the project context (pre: “The smart meeting tool will increase my team’s chances of getting a better grade”; post: “Smart meetings tools increased my team’s chances of getting a better grade.”). UTAUT2 measures were used for the other antecedents of technology acceptance [4]. Effort expectancy is a 4-item scale where higher values indicate ease of use and lower values indicate effort. Items include “I think the smart meeting tool will be easy to use” (pre-survey) and “The smart meetings tools were easy to use” (post-survey). Hedonic motivation is measured using three items including “Using the smart meeting tool will be fun” (pre-survey) and “Using smart meeting tools was fun” (post-survey). Self-efficacy included two items, e.g. “I think I’ll be able to troubleshoot any issues we may

face when using smart meeting tools” (pre-survey). Anxiety included items such as “I feel nervous about using smart meeting tools” (pre-survey). The scale was reverse-coded to reflect the same direction as other technology acceptance antecedents: higher values indicate higher technology acceptance. For the anxiety scale, that means that lower values indicate higher anxiety. Across all technology acceptance antecedents, items were measured on a 7-point Likert scale from 1 = completely disagree to 7 = completely agree.

Language proficiency was measured objectively and subjectively. Test scores, e.g. TOEFL, were used as objective measures. Self and peer assessments were used to measure subjectively perceived language proficiency. Perceived language proficiency is often the more relevant measure because it includes an assessment of how easy it is for a person to communicate with someone else. Formal language proficiency neglects the importance of accents and semantics on the ability to smoothly communicate. Perceived language proficiency is more meaningful for measuring conversational ability, pragmatics norms, and thus a solid working proficiency [6].

Psychological safety was measured using an established 6-scale from Edmondson (1999). Items included “I was afraid of making mistakes” (reverse coded) and “It was easy to ask for help from my team members”. The items were measured on a 7-point Likert scale from 1 = completely disagree to 7 = completely agree.

## 4. Results

Before testing our hypotheses, we checked for homogeneity. For the measured variables, no significant differences were found between groups. Therefore, we did not use a multilevel model to analyze the data.

To test hypotheses H1, we conducted a paired samples T-test (table 1). Speech-to-text technology acceptance changed over the course of the project when teams were using the technology. All antecedents of technology acceptance except self-efficacy showed a significant shift. The direction of the shift, however, differs across variables. Effort expectancy ( $M_{pre} = 4.90$ ;  $M_{post} = 5.17$ ; reverse-coded) and anxiety ( $M_{pre} = 4.80$ ;  $M_{post} = 5.13$ ; reverse-coded) are lower after individuals have used the technology, thus indicating higher levels of technology acceptance. However, performance expectancy ( $M_{pre} = 5.45$ ;  $M_{post} = 4.92$ ) and hedonic motivation ( $M_{pre} = 5.08$ ;  $M_{post} = 4.73$ ) suggest a significant decrease in technology acceptance.

**Table 1. Shift in antecedents of technology acceptance pre- vs. post-project**

	Speech-to-text technology acceptance		<i>T</i>	<i>p</i>
	Pre-project	Post-project		
	<i>M (SD)</i>	<i>M (SD)</i>		
Performance Expectancy	5.45 (1.02)	4.92 (1.35)	-7.36	.00
Effort Expectancy	4.90 (1.27)	5.17 (1.18)	4.18	.00
Hedonic Motivation	5.08 (1.17)	4.73 (1.35)	-5.28	.00
Self-Efficacy	4.86 (1.32)	4.95 (1.31)	1.14	.26
Anxiety	4.80 (1.60)	5.13 (1.53)	3.62	.00

To test H2 and H3, we conducted independent samples T-tests (table 2 and table 3). As hypothesized, non-native speakers display higher levels of speech-to-text technology acceptance antecedents. Differences between non-native (NN) and native (N) speakers are significant on all antecedents of technology acceptance except anxiety. Non-native speakers rate the performance ( $M_{NN} = 5.26$ ;  $M_N = 4.75$ ), ease of use ( $M_{NN} = 5.43$ ;  $M_N = 5.00$ ), hedonic motivation ( $M_{NN} = 5.14$ ;  $M_N = 4.52$ ), and self-efficacy ( $M_{NN} = 5.07$ ;  $M_N = 4.82$ ) higher than native speakers.

**Table 2. Speech-to-text technology acceptance antecedents and language proficiency**

	Language Proficiency		<i>T</i>	<i>p</i>
	Non-native	Native		
	<i>M (SD)</i>	<i>M (SD)</i>		
Performance Expectancy	5.26 (1.16)	4.75 (1.38)	4.06	.00
Effort Expectancy	5.43 (1.12)	5.00 (1.19)	3.77	.00
Hedonic Motivation	5.14 (1.23)	4.52 (1.34)	4.75	.00
Self-Efficacy	5.07 (1.22)	4.82 (1.32)	1.97	.05
Anxiety	5.00 (1.56)	5.11 (1.52)	-0.70	.48

Across levels of psychological safety (H3), only self-efficacy and anxiety show significant differences (table 3). Team members who feel psychologically

safe display higher levels of self-efficacy ( $M_{lowPS} = 4.51$ ;  $M_{highPS} = 4.99$ ) and much lower levels of anxiety ( $M_{lowPS} = 4.43$ ;  $M_{highPS} = 5.13$ ; reverse-coded). Performance expectancy, effort expectancy, and hedonic motivation do not differ across different levels of psychological safety.

**Table 3. Antecedents of speech-to-text technology acceptance and psychological safety**

	Psychological Safety		<i>T</i>	<i>p</i>
	Low	High		
	<i>M (SD)</i>	<i>M (SD)</i>		
Performance Expectancy	4.81 (1.26)	4.95 (1.34)	-0.79	.43
Effort Expectancy	4.95 (1.17)	5.19 (1.19)	-1.56	.12
Hedonic Motivation	4.56 (1.37)	4.77 (1.32)	-1.19	.24
Self-Efficacy	4.51 (1.49)	4.99 (1.23)	-2.41	.02
Anxiety	4.43 (1.58)	5.19 (1.49)	-3.68	.00

In summary, our data widely supports H2. H1 and H3 are partially supported, in the sense that certain antecedents of technology acceptance support the hypotheses while others show either no reactivity to psychological safety (H3) or the opposite shift than hypothesized (H1).

## 5. Discussion

Our results show that speech-to-text technology acceptance changes as the Millennial and Gen Z users in the sample use the technology. Furthermore, team members that are non-native speakers of the team language are more likely to accept speech-to-text technology than their native speaking counterparts. Team members who feel psychologically safe in their team display more self-efficacy and less anxiety with speech-to-text technology.

Our research contributes to the knowledge about technology acceptance, specifically the acceptance of natural language processing technology in global virtual teams. We add language proficiency and psychological safety to the discussion of technology acceptance, and thus expand the reach of the theory to include new elements of user characteristics and team dynamics. We see language proficiency and psychological safety as boundary conditions for technology acceptance and test these in a global virtual team setting. Particularly during the COVID-19 pandemic, virtual teams have become ubiquitous, and we expand the knowledge of their technology use and

its consequences.

To that end, we applied the UTAUT models of technology acceptance to a global virtual team setting. For GVT, we provided empirical evidence that experience with speech-to-text technology, language proficiency of team members, and psychological safety in the team make a difference for the acceptance of automated captions and transcripts.

When comparing the antecedents of technology acceptance at the end of the project – after captions and transcripts had been used by teams – with expectations about the technology before the project started, we find significant shifts for all variables except self-efficacy. The effort of handling the technology and the anxiety towards the new technology is lower after individuals have used the automated captions and transcripts, thus indicating higher acceptance of speech-to-text technology. These findings are in line with our hypothesis that technology acceptance increases with use of the technology and consistent with previous research that found decreasing anxiety and apprehension after using communication technologies [22]. With experience, the technology becomes easier to handle and the anxiety about the unknown eases.

More surprising are the findings that performance ratings and hedonic motivation decrease over the course of the project, representing lower levels of technology acceptance antecedents after use. Millennials and Gen Z expect technology to work seamlessly and without mistakes. While natural language processing has improved significantly over the last few years, automated captions and transcripts are still in a nascent phase with ongoing improvements in accuracy. Transcript accuracy was rated 3.2 (mean) by participant in our study on a scale from 1 (not at all accurate) to 5 (perfectly accurate). In open-ended, qualitative comments on the post-project survey, inaccuracies were attributed to internet connectivity issues, semantic nuances, and people talking over each other, which explains the decreasing performance ratings.

Hedonic motivation likely decreased because participants, particularly Gen Z participants who are at the very beginning of their careers, mostly try new technology in a private setting. While they are usually open towards new technologies and enjoy trying them, workplace technology, such as the technology used in this setting, is less likely to be associated with a hedonic motivation. There may also be a spill-over effect from performance issues. If users don't feel that the technology helps them achieve their goal in a workplace setting, they are less likely to enjoy using it.

The decrease in hedonic motivation may also be caused by what Seeber et al. (2019) call 'devolution of

social interaction' (p. 8) [1]. Norms of social interaction are altered, and the personal element of meetings may be diminished because only task-related content is wanted on the transcript.

Additionally, the lives of all study participants were severely disrupted due to the COVID-19 pandemic while they worked on the virtual team project [33]. Maintaining a functioning team while dealing with private and workplace related anxieties posed a challenge for many team members. Speech-to-text technologies likely created less anxiety and effort than navigating the changing workplace. At the same time, hedonic motivation to use a new technology was low because of cognitive and emotional overload.

Lower performance ratings may also be influenced by how much team members were impacted by the COVID-19 pandemic. The speech-to-text technology could have been a performance driver because team members who had to miss meetings could use the transcripts. However, team member who suddenly had to take care of children at home, worked as first responders in New York City, or faced the emotional burden of a severely sick family member, did not have the time and capacity to use these transcripts.

Generally, the mixed results for different antecedents of technology acceptance are in line with qualitative comments. Sentiment analysis indicates a slightly positive sentiment towards speech-to-text technology (mean = 0.2; with -1 = very negative tone to 1 = very positive tone). The tone of individual comments ranges from -0.8 to 1.

When comparing non-native and native speakers of the team language, our data shows a clear indication that non-native speakers are more likely to accept speech-to-text technology. Anxiety is the only variable that is stable across levels of language proficiency. Because non-native speaker benefit from the reprocessability of written communication [6], it is evident that they'll be likely to rate the performance of captions and transcripts highly. Hedonic motivation is high for the same reason: team meetings are more enjoyable when understanding is less of an issue. Why ease of use and self-efficacy is higher for non-native speakers needs further exploration. Again, spill-over effects are likely part of the explanation. According to qualitative comments, non-native speakers found that the transcript and captions empowered them to be self-sufficient rather than having to ask for clarification and explanation. In this case, the technology helps non-native speakers to gain a more general sense of self-efficacy.

Psychological safety was also hypothesized to make a difference for technology acceptance – with high psychological safety fostering antecedents of technology acceptance. For self-efficacy and anxiety,

we found this hypothesized relationship. It seems evident that team members who feel safe and do not fear negative consequences when they speak up experience lower levels of anxiety in any part of the teamwork, including technology use.

Self-efficacy is higher when an individual feels psychologically safe because that individual can freely access resources in the team to solve issues with captioning and transcription technology. Self-efficacy is a measure of how comfortable an individual feels to troubleshoot problems with the technology. Psychological safety increases learning and the willingness to speak up [32]. So team members who feel psychologically safe will readily reach out to teammates to solve technology issues.

The other antecedents of technology acceptance – performance expectancy, effort expectancy, and hedonic motivation – remain stable across different levels of psychological safety. Particularly for performance expectancy, we would have expected that psychological safety has a positive effect because team members don't need to fear negative consequences of their recorded behavior. However, it appears that users saw performance expectancy as a purely outcome-oriented measure that is unrelated to interpersonal relationships.

## 6. Limitations and directions for further research

While our study contributes to the body of knowledge on speech-to-text technology acceptance in global virtual team settings, some limitations need to be addressed. These limitations can be the basis for future research on natural language processing technology, specifically in a team context.

First, we measured antecedents of technology acceptance at two points in time – before the project began and after it ended. Thus, our data is rather static and does not show the process of accepting or rejecting a technology. Multiple measurements over the course of the project or an observational study promise insights into how and why technology acceptance develops.

Second, performance ratings are potentially influenced by factors not related to the technology, e.g. how hard working team members perceive each other, how much self-confidence someone has in comparison with confidence using the tech.

Lastly, pre-project data was assessed in February 2020, which was before the COVID-19 crisis hit most of the study participants' countries. The post-project survey was administered in April 2020 when the pandemic was causing major disruptions for all study participants. The findings may be impacted by the

COVID-19 pandemic; particularly, pre-/post-comparisons will need to be replicated at a time when pre- and post-living and working situations do not differ vastly.

## 7. Conclusion

This research investigated the acceptance of speech-to-text technology, such as automated captions and transcripts, in global virtual teams. Specifically, this paper addressed whether use of the technology, language proficiency of team members, and psychological safety make a difference for technology acceptance.

Our data showed that while performance expectancy and hedonic motivation are lower after the technology is used, effort and anxiety decrease with use. Non-native speakers display higher levels of technology acceptance. Psychological safety seems to be playing a role for self-efficacy and reduction of anxiety.

Therefore, we conclude that non-native speakers in global virtual teams benefit most clearly from automated captions and transcripts at the moment. Thus, managers should particularly implement speech-to-text technology in teams with non-native speakers. The built-in redundancy of adding a written element, captions, to verbal communication allows them to follow conversations better and ultimately feel more included. Transcripts allow non-native speakers to reprocess content and therefore be better able to contribute to team work. While this study honed in on language proficiency, automated captions and transcripts have the potential to increase inclusivity and accessibility for multiple groups of individuals. Considering that performance ratings decreased after individuals have used the technology, it may be worth reminding all team members, and native speakers in particular, that even though speech-to-text technologies are not perfect, they are worth using because non-native speakers feel empowered and included by them.

The most promising way towards wide-spread acceptance and commercialization of speech-to-text technology remains, however, the continuous improvement of the user interface and output quality. Performance of the technology is an important driver for acceptance, and especially Millennial and Gen Z users expect a seamless experience and highly accurate results.

## 8. References

- [1] I. Seeber, L. Waizenegger, S. Seidel, S. Morana, I. Benbasat and P. B. Lowry, "Collaborating with



- technology-based autonomous agents: Issues and research opportunities," *Internet Research*, vol. 30, no. 1, pp. 1-18, 2020.
- [2] S. C. Srivastava and C. Shalini, "Social presence in virtual world collaboration: An uncertainty reduction perspective using a mixed methods approach," *MIS Quarterly*, vol. 42, no. 3, pp. 779-804, 2018.
- [3] V. Venkatesh, M. G. Morris, G. B. Davis and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425-478, 2003.
- [4] V. Venkatesh, J. Y. Thong and X. Xu, "Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology," *MIS Quarterly*, vol. 36, no. 1, pp. 157-178, 2012.
- [5] H. Tenzer and M. Pudelko, "Media choice in multilingual virtual teams," *Journal of International Business Studies*, vol. 47, no. 4, pp. 427-452, 2016.
- [6] C. Fleischmann, J. Aritz and P. Cardon, "Language Proficiency and Media Richness in Global Virtual Teams: Impacts on Satisfaction, Inclusion and Task Accomplishment," *ACM Transactions on Social Computing*, vol. 2, no. 4, pp. 1-18, 2020.
- [7] J. Elson, D. Derrick and G. Ligon, "Examining Trust and Reliance in Collaborations between Humans and Automated Agents," in *Proceedings of the 51st Hawaii International Conference on System Science*, Big Island, HI, 2018.
- [8] N. McNeese, D. Demir, E. Chiou and G. Yanikian, "Understanding the Role of Trust in Human-Autonomy Teaming," in *Proceedings of the 52nd Hawaii International Conference on System Science*, Maui, HI, 2019.
- [9] A. Hussein, S. Elsayah and H. Abbass, "Towards Trust-Aware Human-Automation Interaction: An Overview of the Potential of Computational Trust Models," in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Maui, HI, 2020.
- [10] I. Seeber, E. Bittner, R. O. Briggs, G.-J. de Vreede, T. de Vreede, D. Druckenmiller, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, G. Schwabe and M. Söllner, "Machines as Teammates: A Collaboration Research Agenda," in *Proceedings of the 51st Hawaii International Conference on System Science*, Big Island, HI, 2018.
- [11] K. Wagner, F. Nimmermann and H. Schramm-Klein, "Is It Human? The Role of Anthropomorphism as a Driver for the Successful Acceptance of Digital Voice Assistants," in *Proceedings of the 52nd Hawaii International Conference on System Science*, Maui, HI, 2019.
- [12] C. Fleischmann, P. Cardon and J. Aritz, "Smart Collaboration in Global Virtual Teams: The Influence of Culture on Technology Acceptance and Communication Effectiveness," in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Maui, HI, 2020.
- [13] N. Wellnhammer, M. Dolata, S. Steigler and G. Schwabe, "Studying with the Help of Digital Tutors: Design Aspects of Conversational Agents that Influence the Learning Process," in *Proceedings of the 53rd Hawaii International Conference on System Science*, 2020.
- [14] F. Hendriks, C. Ou, A. Khodabandeh Amiri and S. Bockting, "The Power of Computer-Mediated Communication Theories in Explaining the Effect of Chatbot Introduction on User Experience," in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Maui, HI, 2020.
- [15] D. Liu, Y. Li and M. Thomas, "A Roadmap for Natural Language Processing Research in Information," in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.
- [16] H. J. Wilson and P. R. Daugherty, "Collaborative Intelligence: Humans and AI Are Joining Forces," *Harvard Business Review*, vol. 96, no. 4, pp. 114-123, July-August 2018.
- [17] M. Chollet, P. Ghate and S. Scherer, "A Generic Platform for Training Social Skills with Adaptive Virtual Agents," in *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, Stockholm, Sweden, 2018.
- [18] C. D. C. Bokhove, "Automated generation of 'good enough' transcripts as a first step to transcription of audio-recorded data," *Methodological Innovations*, pp. 1-14, 2018.
- [19] I. Boughzala, G.-J. de Vreede and M. Limayem, "Team Collaboration in Virtual Worlds: Editorial to the Special Issue," *Journal of the Association for Information Systems*, vol. 13, no. Special Issue, pp. 714-734, 2012.
- [20] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "Extrinsic and intrinsic motivation to use computers in the workplace," *Journal of Applied Social Psychology*, vol. 22, no. 14, pp. 1111-1132, 1992.
- [21] S. Komiak and I. Benbasat, "The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents," *MIS Quarterly*, vol. 30, no. 4, pp. 941-960, 2006.
- [22] J. Aritz, R. Walker and P. W. Cardon, "Media use in virtual teams of varying levels of coordination," *Business and Professional Communication Quarterly*, vol. 81, no. 1, pp. 222-243, 2018.
- [23] M. O'Leary and J. Cummings, "The Spatial, Temporal, and Configurational Characteristics of Geographic Dispersion in Teams," *MIS Quarterly*, vol. 31, no. 3, pp. 433-452, 2007.

- [24] A. C. Fleischmann, J. Aritz and P. Cardon, "Language Proficiency and Media Synchronicity Theory: The Impact of Media Capabilities on Satisfaction and Inclusion in Multilingual Virtual Teams," in *Proceedings of the 52nd Hawaii International Conference on System Science*, Maui, HI, 2019.
- [25] B. Kirkman, J. Cordery, J. Mathieu, B. Rosen and M. Kukenberger, "Global organizational communities of practice: The effects of nationality diversity, psychological safety, and media richness on community performance," *Human Relations*, vol. 66, no. 3, p. 333–362, 2013.
- [26] D. Bhawe, L. Teo and R. Dalal, "Privacy at work: A review and a research agenda for a contested terrain," *Journal of Management*, vol. 46, no. 1, pp. 127-164, 2020.
- [27] A. Edmondson, "Psychological Safety and Learning Behavior in Work Teams," *Administrative Science Quarterly*, vol. 44, no. 4, pp. 350-383, 1999.
- [28] Y. Zhang, Y. Fang, K.-K. Wei and H. Chen, "Exploring the role of psychological safety in promoting the intention to continue sharing knowledge in virtual communities," *International Journal of Information Management*, vol. 30, no. 5, pp. 425-436, 2010.
- [29] C. Gibson and J. Gibbs, "Unpacking the concept of virtuality: the effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation," *Administrative Science Quarterly*, vol. 51, no. 3, p. 451–95, 2006.
- [30] A. Ortega, M. Sánchez-Manzanares, Gil, F. and R. Rico, "Team Learning and Effectiveness in Virtual Project Teams: The Role of Beliefs about Interpersonal Context," *The Spanish Journal of Psychology*, vol. 13, no. 1, pp. 267-276, 2010.
- [31] R. Dibble, L. Henderson and Z. Burns, "The Impact of Students' Cultural Intelligence on Their Psychological Safety in Global Virtual Project Teams," *Journal of Teaching in International Business*, vol. 30, no. 1, pp. 33-56, 2019.
- [32] A. Edmondson and Z. Lei, "Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct," *The Annual Review of Organizational Psychology*, pp. 23-43, 2014.
- [33] P. Cardon, J. Aritz, C. Fleischmann, T. Elhaddaoui, K. Getchell, R. Helens-Hart, K. Iivonen, F. Insignares, X. Li, M. Logemann, J. C. Palmer-Silveira, M. Ruiz-Garrido, A. Shrivastava, S. Springer, J. Stapp, S. Swartz, N. Warren, K. Woolstenhulme and E. Zemaitis, "What Happens to Global Virtual Teams When a Pandemic Hits? Maintaining Normalcy and Stability with Disruption All Around," *The Western ABC Bulletin*, 14 June 2020.