Proceedings of the 5th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'19)
Lisbon, Portugal – August, 2019
Paper No. MHCI 101
DOI: 10.11159/mhci19.101

"Alexa, Open Analyst Assistant": Evaluating a Voice User Interface for Financial Analysis

Azadeh Nematzadeh, Grace Bang, Zhiqiang Ma, Shawn Liu
S&P Global
55 Water, NYC, US
azadeh.nematzadeh@spglobal.com; grace.bang@spglobal.com; zhiqiang.ma@spglobal.com; shawn.liu@spglobal.com

Abstract **-** abstract- Tech-savvy creators have developed countless voice interaction applications for personal and social use, but there is a limited production of voice applications for financial and investment analysis use cases. We show how a voice user interface (VUI) can impact the workflow of financial analysts. Specifically, we hypothesized that a VUI could lead to improvements on financial analysts' job productivity, job performance, and job effectiveness. We ran a controlled experiment where we randomly assigned thirty-two financial analysts to complete six tasks with their current workflow tools, with VUI through the publicly available Alexa skills, or with VUI using an internally developed Alexa skill for financial analysts. Analysts required less time, fewer steps and were able to multi-task for completion of tasks with the use of VUI. We suggest that VUI in the workplace could improve the productivity and ease of financial analysts' workflow.

Keywords**:** Voice user interface, Alexa echo, Intelligent user interaction, Smart work space, Financial analysis, Experiment design.

1. Introduction

The rapid growth of speech pattern recognition techniques, machine learning and hardware platforms have equipped designers with the tools and platforms to develop various voice user interfaces (VUI) for users to more naturally interact with computers or electronic devices.

The market leaders of personal VUI devices such as Amazon's Alexa, Google's Home, and Apple's HomePod are all branding their devices as "smart home device with your voice". Therefore, the designed interactions within these devices tend to lean towards the daily life in the household. Because of this nature, a large body of recent VUI research has concentrated on examining interaction patterns and observations from various personal and social scenarios and their reflections and connections to current HCI theories [13].

The study and application of VUI for work related tasks is less common. VUI researchers have started exploring the application of VUI in the work place [6], and organizations and companies have become interested in using VUI as a virtual assistant. Some companies have built voice interaction skills to handle simple customer interaction tasks and reduce human involvement. For instance, Capital One Bank customers can talk to Alexa to check their balance, track their spending, and to pay bills. Stone Temple Consulting developed a voice interaction skill that answers questions from their customers about digital marketing. However, no voice interaction skill is publicly available for use in the workplace environment, especially in the financial sector. Additionally, the use of VUI and its impact on job productivity, performance, and effectiveness have not been studied.

We study the use of VUI for corporate work activities and in particular, we measure the impact of VUI as a virtual assistant in the workflow of financial analysts at a major financial services company. A financial analyst's responsibilities involve aggregating and analysing numerous data points, making informed assumptions and inferring what the future may hold. The data can come from various sources including financial fillings, news, company provided presentations and third-party sources. A financial analyst then performs financial computations and projections, or data visualization. All of these tasks require significant effort and time.

We form eight hypotheses to determine whether a financial analyst can navigate complex analysis more efficiently and accurately with a VUI device. We measure the possibility of multi-tasking, reduction of the complexity and time of an operation, reduction of an operation's steps, possibility of completing an operation, and the correctness of the answers with the use of VUI compared to currently used methods. If the answers to our hypotheses are validated, VUI can save an analyst

the exploration costs and decrease potential cognitive overload by facilitating exercises including financial computations, data extraction from lengthy unstructured financial documents or data visualization.

For the purpose of this study, we designed and developed a VUI through Amazon's Alexa and the Echo Show device. We designed and ran a controlled experiment with two controlled groups and an experiment group. The participants were financial analysts who work at a major financial services company. The experiment included six tasks that a financial analyst may face regularly regarding information retrieval (including both document retrieval and information extraction), financial computations, data visualization and financial projections.

We quantitatively analysed the data that we collected from the controlled experiment. We found that the VUI assistant offers a more delightful user experience than the traditional workflow. For document and information retrieval operations we show that a VUI assistant can enhance multi-tasking, reduce the complexity, the operation time, and the number of operation's steps. It also enhances multi-tasking and reduces the number of steps in financial computing, visualization, and projections.

This study is one of a few works that scientifically test the impact of VUI in the workplace environment. The findings reported in this work are not limited to the financial sector but can be relevant in any other work place with similar operations. Additionally, this work clarifies some of the design challenges of a virtual assistant through VUI and provides guideline to facilitate the transition from a traditional workflow to a VUI-enabled workflow.

2. Related Work

López et al. conducted a usability and accessibility test on daily life services, e.g. music, agenda, news and weather, provided by voice-based personal assistants from Amazon, Apple, Microsoft and Google [8]. When VUI possess human-like attributes, it can facilitate a conversation in a group setting environment [9, 11, 12]. These findings revealed that human-like traits are critical to future VUI design consideration, because they help people build social attributions with the devices, resulting in more trust and more frequent system use [14].

In the late nineties, voice assistant devices for workspace gained attention by designers who were focused on VUI for customer service, collaborative drawing [5] and teleconferencing [3]. However, voice interaction merely had a peripheral role in those systems mainly because speech recognition technology was not mature until recently.

James and Roelands designed a series of voice interfaces to support navigation within a complex business GUI and discussed the insights of handling efficiency and ambiguity [6]. It has been shown that VUI is able to facilitate meetings in collaborative workplaces more effectively [10]. Kocielnik et al. created a conversational agent using Amazon Alexa to help workers on activity journaling and self-reflecting their work in addition to an already busy chat-based modality [7]. These studies do not indicate similar importance for anthropomorphism as in personal settings. The clear conservational UX distinction between personal ("relational interaction") and business ("productivity-oriented interaction") settings was made in [2].

3. Hypothesis Design

Since we aim to understand the impact of a new technology (VUI through an Alexa skill via an Echo Show device) on analysts' workflow, we followed the technology acceptance model (TAM) [4] while developing our hypothesis. The technology acceptance model (TAM) [1] is a widely-used information system theory which has fundamental implications for human-computer interaction [4]. TAM has been applied in evaluating of user experiences and it can provide a theoretical guideline to conduct usability studies of VUI.

We divided our hypothesis into two constructs of TAM: Perceived Usefulness (PU) and Perceived Ease-Of-Use (PEOU). PU shows the scale to which a person perceives system adoption would enrich her job performance [13, 1]. Our hypotheses consider job productivity, job performance, job effectiveness, and overall usefulness. PEOU shows the scale to which an individual adopting a particular system would be effortless [1]. Our PEOU hypotheses consider how easy or rigid it is to learn the new technology and work with it.

We investigate the interactions of VUI and financial analysts to complete their job-related tasks to determine the impact of voice assistance for employees and, specifically, financial analysts in the financial services industry.

3.1. Perceived Usefulness

- **H1-Job Productivity:** The use of voice interaction increases the likelihood or possibility of a financial analyst multi-tasking in comparison to a traditional interaction.
- **H2-Job Productivity:** The use of voice interaction can decrease the complexity of each financial analyst's task in comparison to a traditional interaction.
- **H3-Job Productivity**: The use of voice interaction requires less time to complete each financial analyst's task in comparison to a traditional interaction.
- **H4 Job Performance**: The use of voice interaction does not impact the ability of financial analysts to complete each task compared to a traditional interaction.
- **H5-Job Effectiveness:** The use of voice interaction does not impact the accuracy of results of each financial analyst's tasks in comparison to a traditional interaction.
- **H6-Overall Usefulness**: The use of voice interaction is more delightful (or preferable) for financial analysts in their tasks in comparison to a traditional interaction.

3.2. Perceived Ease of Use

- **H7:** The use of voice interaction requires a smaller number of steps to complete each financial analyst's task in comparison to a traditional interaction.

We designed a controlled experiment to scientifically test our hypotheses. We defined an evaluation criterion that included more than twenty quantifiable metrics. The data for these measurements was collected during the experiment session and through a post experiment survey.

4. Experiment Design

We designed and ran a controlled experiment with three participant groups which varied in use of a VUI–Amazon Alexa to complete a set of routine tasks. Two groups were the controlled arms and one was an experimental arm. The controls were to measure the effect of not using VUI or using VUI with currently available functionality on financial analysts' performance. The results from both control groups were compared to the results from the group using VUI with internally developed skills for financial analysts to identify the additional improvements.

The experiment consists of six tasks. These tasks remain unchanged for all participant groups. Each task addresses an operation currently completed by financial analysts as part of their workflow. Broadly speaking, a financial analyst's operation typically includes:

- **Document retrieval task:** The analyst gathers multiple and often disparate data sources to consider in her analysis including financial filings, news, market data, earnings call transcripts, sell side equity analyst's opinion, and presentations from company management. All of these documents are located in different sites and require multiple steps, and consequently time, to retrieve.
- **Information retrieval task:** The analyst collects specific data points across various documents depending on the type of analysis she does. For example, for her analysis on a company, she may need information on the risks that a company faces or any new products that a company is launching. The analyst often reads through documents, performs a keyword search in a document or searches in databases to extract the data. Thus, data extraction can be tedious and time consuming, especially for a newer analyst who is not as familiar with a financial reports' structure as a more experienced analyst would be.
- **Financial computation task:** The analyst must then perform financial computations on the numeric data that she gathered to identify meaningful patterns and trends. This can be as simple as basic arithmetic or interlinking the three financial statements and calculating the debt paydown schedule. Analysts often use Microsoft Excel to build financial models and perform computations.
- **Data visualization task:** Analysts use charts and tables to represent the numerous data and comparisons that need to take place. For example, when assessing how the long-term trend of revenue, or any other financial metric, for a company is performing compared to other companies in the sector, an analyst could look at the raw data or visualize the data. An analyst uses Microsoft Excel to draw a graph and visualize data which is not often a time-saving or intuitive way to create a relevant visualization for financial presentations.

- **Prediction/projection task**: Because investment decisions are based on expectations of future performance, an analyst also spends meaningful time performing projections on her forward-looking view of a company. This task combines and builds upon all the other tasks that are listed above for an analyst to then apply her inference on the future state of the company. Analysts could also incorporate top-down factors including macroeconomic data, industry trends, press releases, and alternative data. The ingestion and analysis of all this data requires analysts to complete multiple processes in parallel. They take the potentially infinite amount of data, aggregate it, and provide their expectations of the company.

For the purpose of this experiment, we frame these tasks for the analysis of the publicly traded company Twitter (TWTR). Twitter was one of the most actively traded stocks as of September 7, 2018. Also, Twitter has a diverse archetype of users with regards to age, gender and profession and is a widely recognized company. The following are the defined tasks:

- **Task 1.** What is the stock price of Twitter? Please write your answer below.
- **Task 2.** Find the latest annual report.
- **Task 3.** What are Twitter's business operating segments according to the latest 10K? Please write your answer below.
- **Task 4.** What was the fiscal year 2017 revenue growth rate? Please write your answer below.
- **Task 5.** Graph the historical revenue growth of Twitter. Please upload any files you have used for this task in your folder on the Desktop.
- **Task 6**. Project rolling 12-months revenue growth through June 30, 2018 without referring to any documents released after February 23, 2018. Please explain your reasoning below and upload any files you have used for this task in your folder on the Desktop.

We recruited 32 analysts in the New York office of a financial services company. During recruitment, we did not disclose the purpose of our study to the participants in order to avoid any potential bias from participants while they took part in the study. The participants were instead provided a tangential scenario as follows during recruitment:

The Data Science team is conducting a small study to understand the Corporate financial analytical process. Each experiment had a minimum of 10 participants of different seniority levels and the maximum duration allowed for the experiment was thirty minutes for all six tasks.

Participants of the first control group (No-Alexa) worked on the given tasks in a similar manner as their usual workflow. The second control group (Public-Alexa) was given an Echo Show device and encouraged to use Alexa in conjunction with their regular routines while working on their tasks. Alexa provides publicly available skills and can answer a subset of the finance questions related to data retrieval. An analyst can interact with an Echo Show by triggering Alexa and asking questions to get assistance with their tasks. For instance, an analyst may ask "what is the stock price of Alphabet?"

The experiment group (Internal-Alexa) was given an Echo Show device. They were also given access to the Alexa skill that we developed for internal use for financial analysts to assist them in their tasks. Participants first needed to trigger the internal Alexa skill by saying "Alexa, open Analyst Assistant". From there, they could ask Alexa questions relevant to financial analysis tasks. Participants who used VUI were prompted to interact with the Echo Show device and Alexa before starting the experiment. For instance, an analyst could ask "what is the weather?" Or "what is the revenue growth Alphabet in 2017?" This helped to familiarize participants with the device and its functionality before jumping into the experiment. Prior to conducting the experiment, we performed a pilot test with four participants.

4.1. Experiment session

The study was conducted in an unoccupied closed-door office room. The analysts were provided with technological devices and tools that replicate their conditions in the office. The analysts also had access to the CapitalIQ platform and the internet. Those in the Public-Alexa and Internal-Alexa groups were also provided with the Echo Show device. For all tasks, participants were instructed to assume it was February 23, 2018, the date when the Twitter annual report was publicly released.

Two experimenters attended each experiment session to observe the participant's behaviors and operations and to collect relevant data. All sessions were recorded. We collected data about the following features per task during the experiment session:

- The duration of each task (Duration): We recorded the time, in seconds, that it took to complete a task.

- The number of steps (Steps count): We counted the number of action steps that a participant took to finish at ask. To unify the counting approach among experimenters we defined a counting criteria for Alexa queries, Google searches, spread sheet operations, financial documents, and study documents. For instance, each utterance to Alexa that included asking questions or initiating or closing an interaction with Alexa should be counted as an action step.
- Sentiment: We recorded three sentiments which are happy, frustrated, and neutral. This was to reduce the subjective nature of sentiment observation and simplify the analysis.
- Accuracy of answers (Correctness): For each task, a correct answer was defined and verified with experienced analysts. Participants recorded their answers to each task in a word document and thus we were able to review and verify each answer.
- The workflow of analyst per task: Here we recorded the type of computation and operations that an analyst used to complete a task. These operations can include reading financial filing files, key word searches, or using Excel for computation.
- Finished or skipped:(Completeness) we recorded whether a participant finished a task or skipped it.
- Additional comments: we recorded additional observation.

Each experiment session took a maximum of thirty minutes. The experimenters gave a fifteen minutes warning during the session. At the beginning of the experiment participants were given a consent form to read and sign. Subjects then read through the experiment's description and got started working on the tasks. Participants were asked to write their answer in the study document. We recorded video and audio of the experiment sessions for our future reference. The study description documents end with a link to a survey for the participants to fill out. We explained our survey design approach in section 5. At the end of the experiment we provided a debriefing document to participants to read and inform them about the real nature of the study. We also asked them for additional feedback or comments.

To ensure accuracy of our observational data, we recorded each experiment session and watched and validated the data from the full duration of the session. For instance, some participants jumped from one task to another, which made it difficult to record the metrics during the session.

4.2. Participants

Financial analysts undertake rigorous analytical work to evaluate the underlying performance of a company for investment purposes. This includes tasks such as gathering and analyzing data from reports, company provided documents, and presentations. Analysts also analyze financial statements, forecast future performance and speak with company management and investors.

They perform financial and credit modeling using a combination of Microsoft Excel and internally developed tools. They put together documentation on their analysis and assessment of the companies which get published externally. The variety in methods to complete the tasks can vary greatly by analyst given the lack of consistent approach provided.

Most analysts have a background in finance, accounting, economics or mathematics. Analysts are typically assigned a portfolio of companies in a specific sector such as technology, consumer products, or oil & gas.

Thirty-two financial analysts who are responsible for researching and analyzing U.S. based companies ('Corporates sector') were recruited to be a part of the study. We excluded analysts who were not in the Corporates sector as the tasks for the study were specific to those that a financial analyst in that sector could conduct. The sample includes 11 women and 21 men. There was a relatively even distribution of participants with regards to education level of undergraduate and graduate degrees.

The participants expressed a relatively high level of comfort with technology, for an overall mean value of 7.4 out of 10. More than 55% of participants have used a voice assistance with 28% of participants said that they are using a technology once a week or more.

Participants were randomly assigned to complete all six tasks under one of the three conditions, using their current workflow tools, using the VUI through the publicly available Alexa skills, or using the VUI using an Alexa skill developed for financial analysts. The three conditions did not differ significantly on any of the characteristics measured, including age, education level, years of experience, seniority in the organization, and industry of companies that they cover. There were two analysts who covered companies in the technology sector but they did not cover Twitter directly.

The majority of participants (75%) found the tasks very similar or similar to the tasks that they complete for their work. The results of two participants were excluded from the analysis due to internet connectivity issues that were encountered during their experiment.

4.3. Survey Design

We designed two post-experiment surveys. First, one for the control group that did not have access to the Alexa device. This survey included twenty-one questions. The second survey for the groups that had access to the Alexa device during the study. The second survey had ten additional questions to the first which were specific to the user interaction with Alexa. We collect three types of information thoughout the survey:

- **Participant**: Questions on the backgrounds of participants including age, gender, education level, experience level as an analyst, comfort with technology and ownership of a voice assistance device.
- **User experience**: Questions directly related to the tasks in the experiment that they completed including their self-reported behavior of number of steps to complete a task. Part of the responses to these questions were incorporated into the evaluation of our hypothesis.
- **Willingness to use in the future**: Questions around their willingness and likelihood to interact with and adopt the technology for their analytical workflow. These included their perception of the technological capabilities of the voice assistance device and any privacy or security concerns that they expressed.

4.4. Internal Alexa skill design

Amazon first released the Echo as a stationary device to provide voice interaction and facilitate shopping on the Amazon online store. The Echo uses a cloud-based voice service, Alexa, to communicate with users. Similar to its counterparts, Apple's Siri and Google Now, Alexa uses an "activation phrase" to wake up and then answer a user's command but there has been one big difference between the Echo and Siri or Google Now, the lack of a screen or graphical interface. The Echo Show filled this gap by providing both a voice and graphical interface.

We used the Echo Show device for our experiment to enable both GUI and VUI interactions. Additionally, we developed an Alexa skill, called Analyst Assistant, specialized to perform tasks related to financial analysis. The developed skill includes several use cases relevant to the tasks mentioned above. An analyst can trigger the Analyst Assistant skill to receive help while they are working on financial analysis tasks.

We anticipated potential challenges of interacting with the VUI. To enhance usability, to create a delightful user-experience, and to increase the acceptance of the newly developed product, we followed the process of interaction design in designing user interfaces (UI) while we were designing and developing the Analyst Assistant skill. However, designing a VUI requires further consideration and introduces new challenges in comparison to designing a graphical user interface (GUI). GUI has the benefit of visuals to inform users about its functionalities and guide users on how to interact with it. This important signal is missing in VUI. To tackle this challenge, we designed an introduction message which is stated by Alexa after users trigger the voice interface.

> *WELCOME MESSAGE: "Welcome to the Analyst Assistant skill! You can ask me about a stock price, business operating segments, revenue growth rate, and projected rolling twelve-month revenue growth of companies. I can also find a financial filing or draw a graph. Say help to learn more about my functionality."*

We cannot expect users to remember all the use cases that our system offers. Thus, we designed a help function to assist users, which can be activated by saying 'help' during the interaction. The system will then ask users which specific task they need help with and provide help based on users' answers.

> *HELP MESSAGE: "How can I help you? You can say help with finding business operating segments, help with finding a report, help with revenue growth rate, help with forecast, or help with visualization."*

Thus, our design decision was to instruct users through the help functionality about what type of voice commands they could use and what are the use cases, operations, and tasks the VUI is capable of executing. The help function can suggest a possible utterance when Alexa cannot answer the user's command. For instance, Alexa may tell a user that:

> *HELP MESSAGE:" For instance, you can say Alexa, what was the revenue growth of Twitter in 2017?"*

Another design challenge that we considered was the natural manner of conversation that users may prefer, a human-human conversation, when interacting with a VUI rather than a human-device interaction. We designed several utterances per request to facilitate the user interaction with the VUI which include terms such as 'please', 'can you', or 'do you have'. For instance, the following utterances can be used for task 3. *please tell me what are Twitter's business segments? can you*

tell me what are Twitter's business segments? please tell me what are Twitter's business operating segments according to the latest 10K? what are Twitter's business segments?

We also anticipated that a VUI user may request further information or the user may be curious about the underlying rational behind the VUI's answer, especially when a task relates to looking into multiple data sources, building a financial model, or making a projection. In task 6 we intentionally encouraged users to ask follow up questions. After the answer is given, Analyst Assistant tells the user that she can ask why.

We displayed the text of Analyst Assistant's responses in the Echo Show display. This is particularly helpful since analysts could take time to read and digest the responses. For longer answers, a participant could use the touch screen feature of the Echo Show display screen to scroll the text up or down.

The visualizations that Analyst Assistant creates and the documents that it retrieves are automatically saved in a cloud storage folder. Analyst can easily work with them in their computer or put them in a report.

We ran several pilot tests to ensure that the developed skill was functioning properly and also made enhancement to the user interactions.

4.5. Hypotheses evaluation criteria

To test our hypotheses, we ran two statistical tests for each task to measure differences between the No-Alexa group and the Internal-Alexa group and between observational data from the Public-Alexa group and the Internal-Alexa group.

Table 1: Data source and metrics to test each hypothesis. S means that the data input comes from a survey and O means that the data input is collected during the experiment session.

	H1	H2	H3	H4	H5	H6	H7
Multi-task	S						
Complexity		S					
Duration			S-O				
Completeness				O			
Correctness					S-O		
Sentiment						O	
Steps count							S-O

We used recorded data from the post-experiment survey and observational data from the experiment session. Table 1 shows the metrics in rows and hypotheses in columns. If a metric was used to test a hypothesis, we marked the cell with its data source. We use S to show that the data input is collected from the post-experiment survey. We use O to show that the data input is observed and collected during the experiment session. For some hypotheses, for instance H3, we collecte data from both the survey and during the experiment session. This is due to the nature of the data we collected, to avoid potential biases or errors.

Table 2: Wilcoxon statistics p-value of hypothesis 1 for each comparison per task.

	task1	task2	task3	task4	task5	task6
Internal-Alexa vs No-Alexa	0.153	0.192	0.03	0.051	0.005	0.005
Internal-Alexa vs Public-Alexa	0.086	0.03	0.004	0.235	0.005	0.044

To test the hypotheses, we compute the Wilcoxon signed-rank statistic between both groups (controlled and experiment) for H1 to H7 except H5. For hypothesis five, we compute the fraction of each sentiment at each group per task.

5. Results

5.1. Multi-tasking

We tested if multi-tasking is more likely to be perceived as easier during voice interaction than during traditional interaction. We collected the data from a survey question in which we asked participants if they were able to multi-task while working on each task, by which we mean they were able to work on two or more process in parallel. (see Table 2) We reject the null hypothesis when comparing the Internal-Alexa group with the No-Alexa group for complex tasks of the experiment which were information retrieval, financial computation, visualization, and projection. Thus, a VUI can facilitate multi-

tasking when analysts are working on a complex task. However, since easier tasks do not require two or more parallel steps we did not get significantly different statistics for task 1 and task 2. We reject the null hypothesis for all tasks but financial computing when comparing the Internal-Alexa group with the Public-Alexa group.

5.2. Complexity

We tested if the complexity of financial analysis operations would decrease when using a VUI. We reject the null hypotheses for the document retrieval, information retrieval, and projection task between the Internal-Alexa group and the No-Alexa group. We reject the null hypotheses for the document retrieval and information retrieval task between the Internal-Alexa group and the Public-Alexa group. (see Table 3)

These findings were quite surprising. We were expecting to accept the null hypothesis for task 4, task 5, and task 6. Participants were able to successfully navigate the VUI for the complex tasks to simplify the multitude of steps normally required. In survey responses, participants indicate tasks 4-6 to be the most complex. Participants likely answered this survey question based on their pre-defined notion of the complexity of the task regardless of how they got their answer for the tasks.

Table 3: Wilcoxon statistics p-value of hypothesis 2 for each comparison per task.

	task1	task2	task3	task4	task5	task6
Internal-Alexa vs No-Alexa	0.059	0.005	0.184	0.32	0.22	0.018
Internal-Alexa vs Public-Alexa	0.059	0.028	0.505	0.553	0.534	0.151

5.3. Time Spent

We tested the hypothesis to understand if it takes less time to complete each financial analyst's task during voice interaction than during traditional interaction. We used time duration as a metric to test this hypothesis. No participant exceeded the thirty-minute time limit(see Table 4).

We reject the null hypothesis when comparing the Internal-Alexa group with the No-Alexa group for task 1, task 3, task 5, and task 6. Task 2 was a simple document retrieval in which analysts were expected to pull out the financial filling of Twitter. Since analysts routinely work with the financial filling reports they know the web site to retrieve the filings. Thus, we did not find a significant difference between using Alexa and not. However, this may not hold if analysts were asked to retrieve other documents. For instance, two participants in the Internal-Alexa group answered in a survey question "What other work-related tasks would you use Alexa for?" with, "Retrieving documents" and "Company news".

We also did not find statistical significance for task 4 when comparing the Internal-Alexa group with the No-Alexa group. Task 4 is about financial computing and participants were asked to compute the revenue growth rate for the fiscal year 2017. This can be due to the fact that analysts may already have the input for the computation of task 4 ready from their work on the previous task. Thus, it did not take them that long to compute this value.

When comparing the Internal-Alexa group and the Public-Alexa group we reject the null hypothesis for all tasks except task 1. This result was expected since task 1 was pretty simple and publicly available skills in Alexa can answer it.

Table 4: Wilcoxon statistics p-value of hypothesis 3 for each comparison per task.

	task1	task2	task3	task4	task5	task6
Internal-Alexa vs No-Alexa	0.014	0.445	0.005	0.169	0.074	0.007
Internal-Alexa vs Public-Alexa	0.102	0.037	0.007	0.012	0.04	0.016

5.4. Task Completion

We tested the hypothesis to evaluate whether voice interaction impacts the completion of each financial analyst's task compared to traditional interaction. We measured whether the task was finished by the participant. We programmatically identified tasks where there was no variation in completion of tasks and thus did not run the Wilcoxon test for those tasks(see Table 5).

The use of a voice interaction devices does not impact the performance of a financial analyst's ability to complete each task. Participants were able to adapt to the VUI and selectively use voice interactions for scenarios that were most suitable

for completion of the tasks. For task 5, two participants did not complete the visualization in the Public-Alexa and Internal-Alexa group. For task 6, two participants did not complete the projections task in the No-Alexa and Internal-Alexa group. While these results are not statistically significant, both visualization and projections are cognitively more complex tasks and thus given the lack of familiarity with the company and the time pressure, the participants may have chosen to skip the tasks.

Table 5: Wilcoxon statistics p-value of hypothesis 4 for each comparison per task.

	task5	task6
Internal-Alexa vs No-Alexa	0.37	1.0
Internal-Alexa vs Public-Alexa	1.0	0.37

5.5. Correctness

We tested the accuracy of results for each financial analysts' task when they use a VUI in comparison to a traditional interaction. We used survey data and correctness metrics to test this hypothesis. We programmatically identified tasks where there was no variation in completion of tasks and did not run the Wilcoxon test for those tasks(see Table 6). The VUI results are as accurate as the traditional results for each financial analyst's task. For task 3, three participants in the control group and one participant in the Public-Alexa group were incorrect in their response, potentially due to their lack of familiarity with identifying the business segments for technology companies. For task 4, one participant in the Public-Alexa group entered the revenue growth rate for the incorrect time frame. For task 5, three participants in the control group, two participants in the Public-Alexa group and two participants in the Internal-Alexa group did not produce a graph or incorrectly graphed the data. For task 6, 5 participants in the control group, 5 participants in the Public-Alexa group and 6 participants in the Internal-Alexa group did not correctly project and explain their rationale for the projections. Again, task 6 is more cognitively complex and requires more time and proficiency of the company and sector.

In the survey, we proceeded to ask participants in the Public-Alexa and Internal-Alexa groups whether they believed the responses provided by the VUI were correct for each task. For all tasks except for Task 1, there was statistical significance in the degree to which participants believe the VUI produced the desired result. When Alexa failed to understand the participant's request in the early stages of the participant's use of the technology, they were unlikely to revert back to the VUI. However, when Alexa failed to understand the request after 3 or more successful returns, participants were more likely to persist in retrieving the correct data from Alexa (see Table 7).

Table 6: Wilcoxon statistics p-value of hypothesis 5 for each comparison per task.

	task3	task4	task5	task6
Internal-Alexa vs No-Alexa	0.083	N/A	0.655	0.705
Internal-Alexa vs Public-Alexa	0.317	0.317	1.0	0.705

Table 7: Wilcoxon statistics p-value of hypothesis 5 for each comparison per task.

	task1	task2	task3	task4	task5	task6
Internal-Alexa vs Public-Alexa	0.317	0.002	0.002	0.019	0.003	0.005

5.6. Delightfulness

We tested the delightfulness of the users' experience of results for each financial analysts' task when they used a VUI in comparison to a traditional interaction. We observed and recorded the participants' sentiment (happy, frustrated, and neutral) during the experiment session.

After running all the experiments, we counted the total number of each sentiment in each group per task. Here, we compare the overall sentiment of the experiment group with the two controlled groups (see Figure 1). Y-axis shows the the total number of each sentiment and x-axis shows the tasks.

For each task we showed the sentiment of each study group with a bar chart. We color coded the total number of sentiments at each group.

We found participants in the Internal-Alexa group were happier than both controlled groups during all tasks.

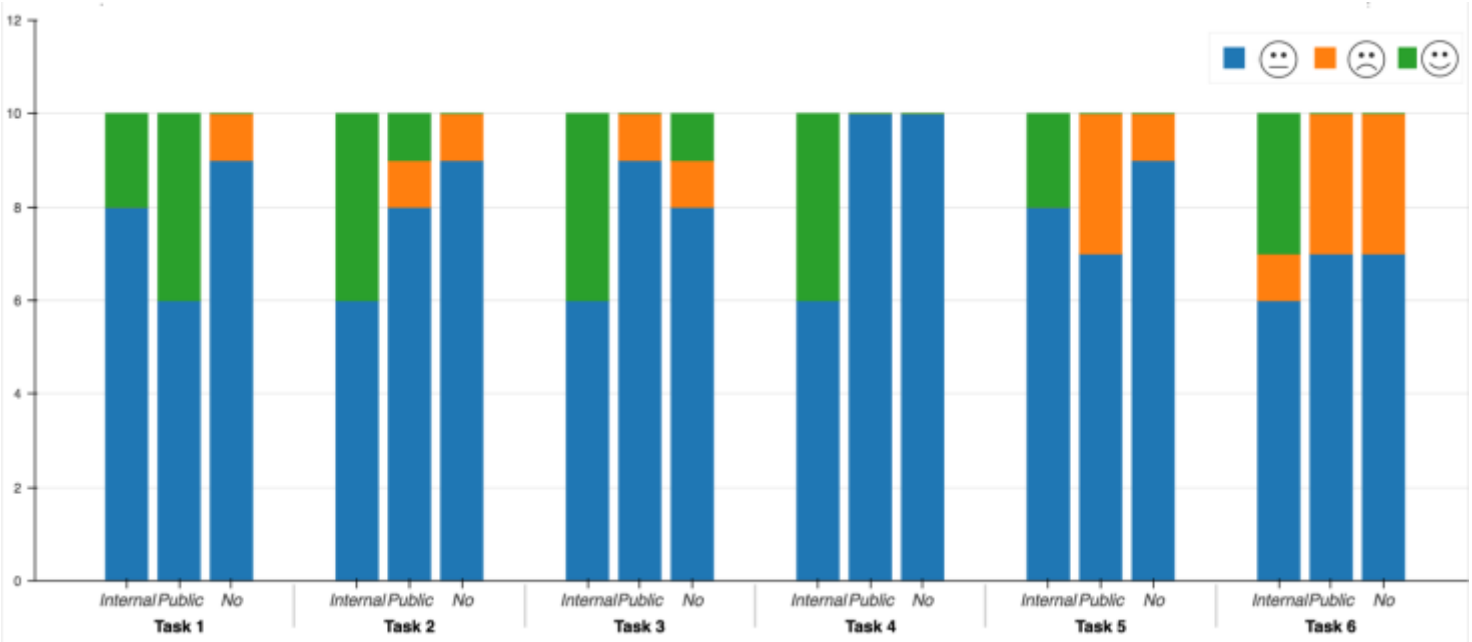

Fig. 1: Sentiment of participants of each group.y needs label.

5.7. Ease of Use

We tested the ease of use for a VUI with regards to the number of steps necessary to complete each financial analysis task in comparison with the traditional approach (see Table 8). We reject the null hypothesis when comparing the Internal-Alexa group with the No-Alexa group for all the tasks except task 2. We reject the null hypothesis when comparing the Internal-Alexa group with the Public-Alexa group for all the tasks except task 1. We did not expect to find a statistically significant p-value for task 1 since Public-Alexa skill has a similar functionality as Internal-Alexa.

Table 8: Wilcoxon statistics p-value of hypothesis 7 for each comparison per task.

	task1	task2	task3	task4	task5	task6
Internal-Alexa vs No-Alexa	0.032	0.138	0.00	0.00	0.00	0.00
Internal-Alexa vs Public-Alexa	1.0	0.027	0.11	0.016	0.00	0.00

5.8. Security considerations

The survey asked two questions to participants in the Public-Alexa and Internal-Alexa condition on security concerns with utilization of a VUI device in the office. Only 18% of the participants stated high levels of security concern with having the VUI in the office. Half of the participants expressed interest in having a VUI on the work desk. Given the confidential nature of the data that financial analysts work with, the lack of security by the majority of participants was surprising.

5.9. Qualitative observation

The survey asked participants in the Public-Alexa and Internal-Alexa groups what other work-related tasks they would use Alexa for. Three out of ten participants in the Public-Alexa answered this question and mainly described currently available Alexa skills such as scheduling meetings on the calendar, calling, and conducting email setup.

All ten participants in the Internal-Alexa group answered this question. They mentioned utilizing Alexa for routine tasks such as finding an email or setting up calendar appointments. Others mentioned similar tasks as those completed in the study such as looking up industry data, conducting extracted data from press release and prior financial reports, and retrieving and downloading other relevant documents and news. Some participants asked for more cognitive complex and inference-based tasks relevant to their job, including finding competitors, conducting peer comparisons, summarizing earnings transcripts and aggregating news.

6. Discussion

The experiments provide evidence of increased job performance, productivity, and effectiveness with the use of voice interaction compared to traditional interactions. Financial analysts who used the VUI were able to complete data and document retrieval, financial computation, visualization and projection tasks in a significantly reduce time frame with less effort required. The VUI also reduced complexity of data and document retrieval tasks. The VUI was perceived as useful for all tasks except financial computation tasks because it requires much fewer steps to change computations ad hoc than through voice. In addition, there was no impact on the performance of the tasks with regards to accuracy or completeness with the use of VUI compared to the traditional methods of interaction. Analysts overall found the VUI for financial analysts to be enjoyable and the tasks more pleasurable. Here we elaborate on other insight and observations beyond our hypothesis framework as follows:

Noise concerns: While the study was conducted in a closed-door experiment room, most offices have open floor plans. In fact, one of the questions that participants consistently asked was the practical application of this technology in an open floor office. A headset can mitigate most of the noise concerns but voice communication to the device could be still disruptive.

Gaining users' trust: We also focused on understanding if participants trust the Analyst Assistant VUI with their tasks. We did not observe any strong signals of distrust. A few participants wanted to know more on how the Analyst Assistant VUI comes up with the answers in the projection task. One participant mentioned that if she were to be given the Analyst Assistant VUI for her work, she would derive the answers in parallel for the first few times of interaction to confirm the answers.

Usability: Through observation, we received several valuable insights on analyst's interaction with the VUI. In the Public-Alexa group one participant did not use Alexa at all while working on his tasks. Nine participants used Alexa for the first task and they received an answer. Five participants used Alexa for the second task but they did not receive the answer from Alexa. Only three participant used Alexa for the third task and did not receive an answer but the fourth task was answered. One participant used Alexa for task 5 and did not receive an answer. No participants used Alexa for task 6. From the feedback we received from the participants after debriefing them, we learned that some participants were not fully aware of what Alexa could do. For instance, participants could have used Public-Alexa instead of reading a financial filings report to find the data input for the visualization task but none of them did. Also, some participants forgot to use Alexa or gave up after they did not get the answer for question 2. Other comments we received after the debrief were that some participants were biased toward their past experiences with using a VUI. One candidate mentioned that she used Apple Siri before and Siri was not recognizing her accent. A few candidates said that VUIs often provide irrelevant answers or they provide extra information.

- **User archetypes:** Alexa can be more useful for junior analysts, especially because they are not as fast and as familiar with workflows and reports as senior analysts. The VUI assistant can inform an analyst about the rationale behind the scores, and show her various financial filings that are required for the computation.
- **Transition:** This study gives us insight on ways to introduce new technological tools in the workplace environment and on what facilitates employees to continue their use of the technology. We believe that it is important for analysts to be able to combine the outputs from new technologies with their intellectual overlay.

7. Conclusion

The findings of this study have broad implications for the use of VUI in the workplace environment. When a VUI is properly designed and users are aware of the full capabilities of this technology, users are delighted to use voice interaction with computers. VUI facilitates a more natural and 'human' way of interaction.

Future work can be extended to observe the natural interaction of financial analysts with VUI in their workplace environment and with a broader range of tasks. We could extend the study to a larger participant base and analyze differences in analyst behavior with VUI for other attributes of the participants such as seniority level, comfort with technology, and the sector or industry that they cover.

References

[1] Fred D. Davis. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3):319, sep 1989.

[2] A Følstad and M Skjuve. Business and pleasure? Relational interaction in conversational UX. In The CHI' 18 Workshop on Voice-based Conversational UX Studies and Design, 2018.

[3] David Frohlich and Owen Daly-Jones. Voicefax: a shared workspace for voicemail partners. In Conference companion on Human factors in computing systems - CHI '95, pages 308–309, New York, New York, USA, 1995. ACM Press.

[4] Kasper Hornbæk and Morten Hertzum. Technology Acceptance and User Experience: A Review of the Experiential Component in HCI. ACM Transactions on Computer-Human Interaction, 24(5):1–30, oct 2017.

[5] H. Ishii and H. TeamWorkStation: towards a seamless shared workspace. In Proceedings of the 1990 ACM conference on Computer-supported cooperative work - CSCW '90, pages 13–26, New York, New York, USA, 1990. ACM Press.

[6] F James, J Roelands Proceedings of the fifth international ACM, and Undefined 2002. Voice over Workplace (VoWP): voice navigation in a complex business GUI. In Proceedings of the fifth international ACM conference on Assistive technologies - Assets '02, pages 197–204, 2002.

[7] K. Author. (2015, May 10). Facility Greenhouse Gas Reporting (2nd ed.) [Online]. Available: http://www.ec.gc.ca/ges-ghg/default.asp?lang=En&n=040E378D-1, g and prediction theory," J. Basic Eng., vol. 83, no. 4, pp. 95-108, 1961.

[8] Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. Designing for Workplace Reflection: A Chat and Voice-Based Conversational Agent. In Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18, pages 881–894, New York, New York, USA, 2018. ACM Press.

[9] Gustavo López, Luis Quesada, and Luis A. Guerrero. Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. In International Conference on Applied Human Factors and Ergonomics, pages 241–250. Springer, Cham, 2018.

[10] Ewa Luger and Abigail Sellen. "Like Having a Really bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16, pages 5286–5297, New York, New York, USA, 2016. ACM Press.

[11] M McGregor, JC Tang CSCW, and Undefined 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17, pages 2208–2220, 2017.

[12] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. Voice Interfaces in Everyday Life. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, pages 1–12, New York, New York, USA, 2018. ACM Press.

[13] Martin Porcheron, Joel E. Fischer, and Sarah Sharples. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17, pages 207–219, New York, New York, USA, 2017. ACM Press.

[14] François Portet, Michel Vacher, Caroline Golanski, Camille Roux, and Brigitte Meillon. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. Personal and Ubiquitous Computing, 17(1):127–144, jan 2013.

[15] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18, pages 857–868, New York, New York, USA, 2018. ACM Press.