Department of Business and Management.
Thesis on Customer Intelligence & Big Data.

Alexa! A Study Concerning User's Acceptance of Recommender Systems from Smart Speakers

SUPERVISOR
Prof. Giuseppe
Francesco Italiano

CANDIDATE
Matteo Cennamo
Mat. 706241

CO-SUPERVISOR
Prof. Alessio Maria
Braccini

Academic Year
2019 - 2020

Alexa! A Study Concerning User's Acceptance of Recommender Systems from Smart Speakers

Matteo Cennamo
LUISS Guido Carli
- a.y. 2019/2020 -

SUMMARY

1. INTRODUCTION

The collective imagination has been attracted by machines which behave in human-like manner since the beginnings of the XX century, when the scientific fiction was born together to the concept of “robot” that is still today a fascinating field of research among the scientists and the object of movies and novels. Even at dawn of the computer science, the people thought of how they would feel interacting with humanoids, while the scientists attempted to resolve the practical issues implicated in the development of such devices.

Leaving apart the mechanical movements, the responses to the environment stimuli and other interactions, there is a capability which has considered for a lot of decades a prerogative of human thought and mind, I mean the capability to understand and speak the natural language. Indeed, many animals that show impressive language skills are considered “more intelligent” than the others, namely dolphins, wales, chimpanzees. Some species of parrots and crows are able to replicate human speech and can learn more than fifty sentences. However, it was believed that the reproduction and the understanding of language was such a peculiar characteristic of the human being that it was associated with the ability to develop thought and consciousness. This was the idea of the patron of computing science Alan Turing when designed the “imitation game”[1]. To answer the question “Can machines think?”, Turing design a hypothetical test based on one human interrogator and two interviewees, a man and a machine. Each agent occupies a different room, and the interrogator can ask to the interviewees the question he/she likes to understand who is the man and who is the machine. In order that tones of voices do not affect the interrogator’s judgement, the answers should be typewritten. Turing quotes that the machine can “think” if the answers it provides are indistinguishable from the man’s ones, meaning that the interrogator can not recognize who of the two interviewees is the machine.

To give an example of how far the technology has gone in the last seventy years, we can cite Professor Jefferson's Lister Oration for 1949, who said: “Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt”[2], which even nowadays seems a pretty reasonable statement and many people agree with it. However, in 2019 an artificial intelligence developed by Huawei completed the Schumbert’s No 8 symphony, also known as the “Unfinished Symphony”. The task was accomplished by a dual neural processing unit (NPU) in the Huawei Mate

[1] Turing, A. M. (2009)
[2] Jefferson, G. (1949).

20 pro smartphone[3]. It was a great publicity stunt, with the purpose of show the capabilities of the new smartphone, but it astonished the public because of the common belief that a robot can't make artistic products. Likewise, even more AIs are involved in design tasks, which is known to require fancy and imagination. In 2017, Ferrero launched a project called "Nutella Unica", which involves the employment of an AI to create around seven million of unique packages for the famous nut and cocoa cream, relying on a database of patterns and colours[4]. Here, we can state that it is impossible to say what the next generation of AI could be really able to do.

Nowadays both the calculation and the storage capacities of the computers experienced exponential growth, while the improvement of the techniques in the AI field allowed the emergence of always better algorithms of voice recognition. Hence they achieved appropriated levels of efficacy, many tech companies started the production of devices endowed with such algorithms. Google launched the Voice Search app for smartphones in the second half of the 2000s, then, three years later, Apple introduced Siri, which is now a prominent voice recognition assistant, while Amazon developed Amazon Alexa and Microsoft's Cortana.[5] Amazon launched the first smart speaker (Amazon Echo) in 2014 becoming the leader in the market. A smart speaker is a table device connected to the Internet and to other devices (e.g. smartphones, smart plugs, etc.). The user can interact to that device through voice commands and receive the answer to questions or can performs some actions communicating with other devices, such as shutting down the TV, switching off the light or setting an alarm. In five years, the smart speaker market experienced exponential growth, especially in U.S., where 66.4 million of households use smart speakers in 2019 and Amazon remains the market leader with 61.1% of market share, followed by Google (23.9%).[6] In the world, 38.5 millions of smart speakers were been sold in 2018[7].

In this period of expansion, it is fundamental for the marketers to understand the customers' needs, how they use these devices and how to extract value from the users. My work takes place in this background.

1.1. THE SMART SPEAKERS

The smart speakers are table devices with an internet connection and equipped with voice assistant software. Thus, the smart speakers are always connected to the Internet, since they need a cloud-

[3] Kennedy, John (2019).
[4] Oh, Jasmine (2019).
[5] TechTarget (2018, January).
[6] Kinsella, B. (2019 [3]).
[7] Kinsella, B. (2019 [2]).

based voice assistant, such as Amazon Alexa, Apple Siri or Microsoft Cortana. They are able to perform some tasks interacting with the user through voice commands. It was stated by Edu *et al.* (2019)[8] that "voice is one of the most effective and expressive communication tools" and it differs from the usual human-machine interactions, which are based on screen (or tactile screens), mouse and keyboards. Voice commands can improve user experience and simplify some routinely tasks, even for people who are not confident with the technology, thanks to a human-like interaction.

The smart speakers take advantage and are been developed from a combination of techniques and recent technologies:

1) **Artificial Intelligence** (AI) and **machine learning**: this is a field of computer science that deals with the issue of building algorithms that can react according to environmental stimuli. It could be described as the endeavour to replicate human intelligence in machines. The machine learning consists on endowing the algorithms with a flexible apparatus which is able to "learn" from past experiences and behave accordingly to them.
2) **Natural Language Processing** (NLP): this field involves the methodologies of natural language analysis, processing and structuring. The "natural languages" are all the languages commonly spoken by human-beings, such as English, Spanish or Japanese. They are opposed to "artificial languages", which are the machine or arithmetical ones.
3) **Cloud computing**: this is a set of technologies and infrastructures which provide the user to a on-demand computing service. Storage and computing capabilities of the single device could be enhanced, delegating the tasks that are too computationally expansive for the device to a cloud service provider who owns the data centres and the computing infrastructure.

To work, the smart speakers need a sophisticated infrastructure made up: the physical device the user interact with; a personal voice assistant which is able to comprehend and speech the natural language; a cloud service provider that delegates the commands to other software applications; other connected devices which can be controlled by the smart speaker. The architecture underlies the Smart speaker usage is shown in Figure 1[9].

[8] Edu, J. S., Such, J. M., & Suarez-Tangil, G. (2019).
[9] Abdi, N., Ramokapane, K. M., & Such, J. M. (2019).

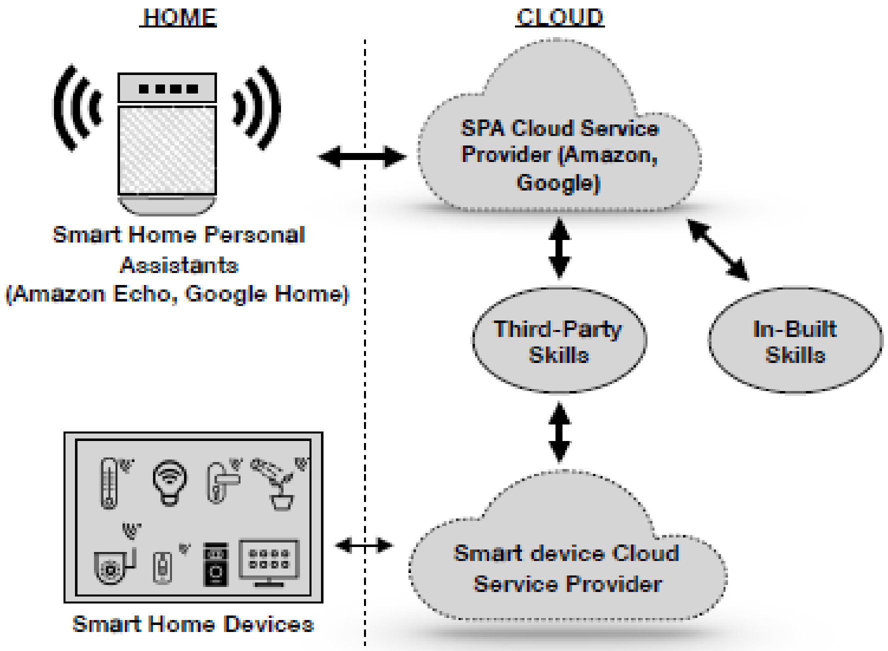

Figure 1: Smart speakers underlying architecture and its key components. Source: Abdi et al. *(2019).*

The user interacts with the smart speaker (the "smart home personal assistant" in the Figure 1) via voice commands. Then, the command is processed remotely by a SPA (Smart home Personal Assistant) cloud service, such as Amazon or Google. Once the command is interpreted, the cloud service provider delegates the requests to a set of "skills" the command refers to. A "skill"[10] is a software application that provide the answer to the users' commands. The skills are usually distinguished in two branches: *in-built* and *third-party* ones. In-built skills are embedded in the smart speaker by the manufacturer of the device and are used to settle alarms, ask the hour and they accomplish other default commands. Whereas, third-party skills are created by other marketers and involves the interrelationship with other smart devices, such as smart bulbs, smart TVs, smart plugs, and so forth. Therefore, the smart speaker represents only the interface with the user of an architecture with not always clear boundaries, which involves many service providers and the connection with many devices. The user can even develop his/her *custom skills*, using web services like Tasker or IFTTT (If This Then That)[11].

[10] We adopt Amazon's nomenclature because it is the market leader, however the other marketers use different terms, such as "actions" for Google's devices.

[11] Hoy, M. B. (2018).

The number of skills sharply rises in the last years and Amazon still remains the leader, having the major number of skills embedded in its devices. In September 2019, Amazon's Alexa reached 100 thousand of skills all over the world, though not all the skills are available everywhere. U.S. counts more than 65,000 skills, while Italy does not reach 3,000[12]. Figure 2 shows the number of Alexa's available skills by country in September 2019.

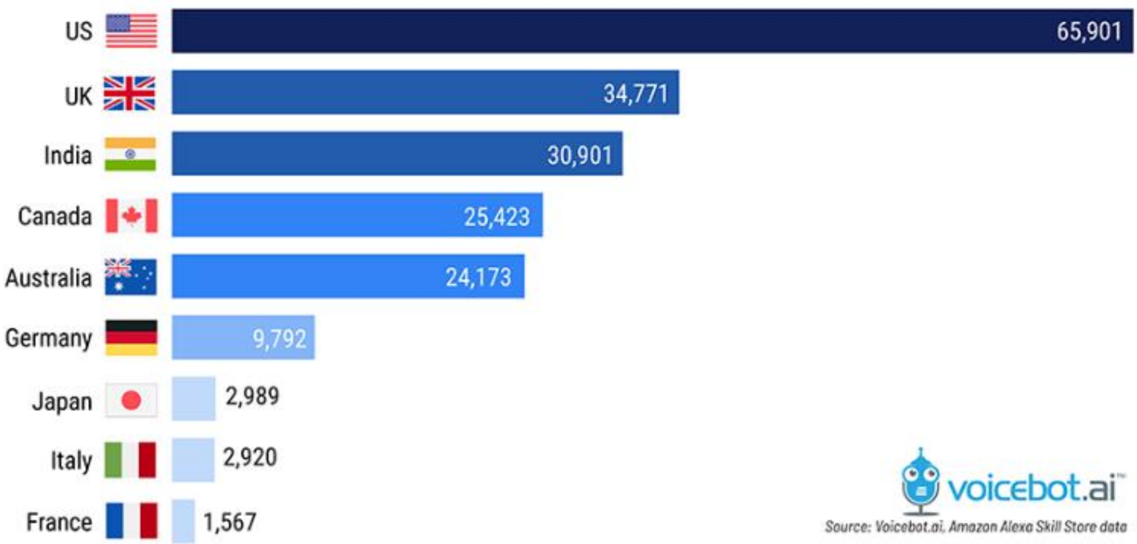

Figure 2: Alexa's number of available skills by country. Source: Voicebot.ai, Amazon Alexa Skill Store Data (updated: September 2019).

Kinsella[13] reported that in January 2019 the number of "actions" ("skills", in Amazon's nomenclature) increased of 2.5 times with respect to the previous year, reaching 4,253 in U.S.

The smart speaker needs to be always switched on in order to receive the commands. However, the device is activated only when a key word is uttered by the user. The default key word for Amazon's devices is "Alexa!", while for Google's ones is "Hey, Google!", though it could be customized by the user. Once activated, the smart speaker informs the user whether the command was successfully or unsuccessfully completed by the device.

Form 2014, when Amazon Echo (the first smart speaker) was released, many models and brands attempted to grasp a slice of the market. Google Home was launched in November 2016, whereas the explosion of the market, by marketers' side, occurred in 2017. On May 2017 the Amazon Echo Show was launched, the first smart speaker equipped with a tactile LCD screen; on September of the same year, Echo Spot was released, which accounts similar functionalities of Amazon Echo previous version, but with an hemispherical look; Echo Dot was released on October 2017, a cheap device that

[12] Kinsella, B. (2019 [4]).
[13] Kinsella, B. (2019 [1]).

is able to plug into other smart speakers[14]; on the same month, Google Home Mini was released, a smaller version of its precursor, together with the Sonos One, which is more like a smart stereo controlled by voice commands and with Alexa embedded in (but it is even compatible with Google assistant); On December, Home Max was launched, a device with larger stereo speakers than the previous versions of Google Home. Apple devices waited February 2018 to launch its first smart speaker, Apple HomePod, embedded with Siri voice assistant. On September 2018 Amazon released an updated model of Echo, calling it Amazon Echo Plus; Google replied on October, launching Google Home Hub. In the years, many versions of these devices have been sold, namely the Amazon Echo reached the third generation. On January 2019, these smart speakers make up the 90% of U.S. market[15].

The Table 1 below exhibited the main releases in chronological order.

DATE	BRAND AND NAME	FEATURES
November 2014	Amazon Echo	First smart speaker in the market.
November 2016	Google Home	First Google’s smart speaker. Possibility to purchase products by voice commands.
May 2017	Amazon Echo Show	Equipped with tactile LCD screen.
September 2017	Amazon Echo Spot	Equipped with a screen and similar to a hemispherical-look alarm.
October 2017	Amazon Echo Dot	Small and cheap device able to plug into other smart speakers.
October 2017	Google Home Mini	Smaller version of Home.
October 2017	Sonos One	Good quality stereo speakers for the music lovers.
December 2017	Google Home Max	Bigger stereo speakers then Home.
February 2018	Apple HomePod	High audio quality and works with Apple Music. Smart home hub functionalities.
September 2018	Amazon Echo Plus	Similar to Echo, but with smart home hub functionalities.
October 2018	Google Home Hub	Similar to Home, but with smart home hub functionalities.

Table 1: The development of the market of smart speakers. Release date and special features of the smart speakers are provided in the table.

[14] Smith, K. T. (2018).

[15] *Voicebot* (2019, January). Consumer Smart Speaker Consumer Adoption Report. *Voicebot.ai*.

1.2. THE ROLE OF SMART SPEAKERS IN SMART HOMES

Common tasks that smart speakers can accomplish are setting an alarm, create a shopping list, update the calendar, play a music, but even interact with other connected smart devices to turning on the light, setting the thermostats, switch on the TVs. For these reasons, it has been stated that smart speakers can became the central hub of a new generation of smart homes. Gram-Hanssen & Darby (2018) said that “There is no fixed definition of a smart home, but an understanding that smart homes incorporate digital sensing and communication devices”[16]. Smart homes are projected to simplify daily tasks and, at the same while, to save energy and time. The IoT devices constitute the atomic elements of smart homes and they rely on a wide range of sensors that capture the state of the environment. Then, the data gathered by the sensors are collected and analysed by a controller. According to the state of the environment, the controller decides whether activate or not the actuators, which are physically implied on changing the environmental conditions (the actuators are pumps, valves, relays, and so forth). Sensors, controllers and actuators constitute the so-called *control systems*. An example of an IoT device is the smart air conditioner, which is a control system that works as it is shown in Figure 3.

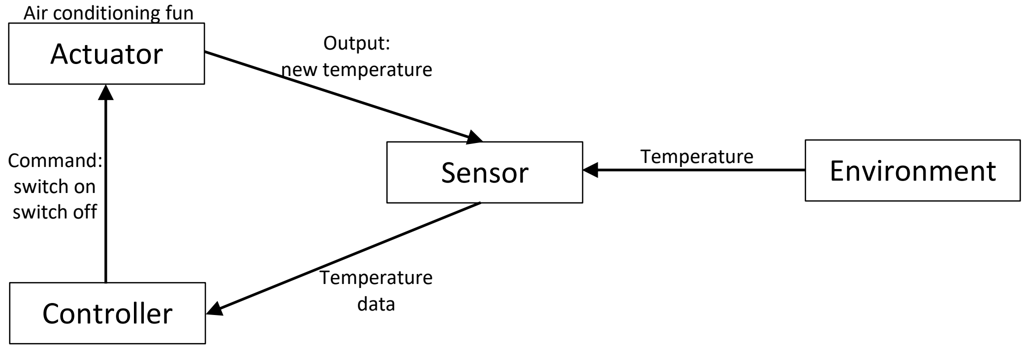

Figure 3: control system of a smart air conditioner. The sensor measures the temperature of the environment and sends the data to the controller. The controller is embedded with an algorithm and decides whether switch on or switch of the actuator (the fun). The output of the control system is a new environmental temperature which is again measured by the sensor to start a new loop.

The controller of an IoT device, sometime, is not powerful enough to analyse the data gathered by the sensors, therefore delegates a cloud service provider the computations or the storage capacities needed to accomplish the tasks. This is the reason why the smart speakers are connected to a cloud service provider. Since the user want a quick response by his/her smart speaker, these devices stored the dictionaries (or “lexicons”) in their RAM, which is a rapid access memory, though it is usually not so large. The combination of these two solutions endows the smart speakers with an agile data analysis and processing.

[16] Gram-Hanssen, K., & Darby, S. J. (2018).

The smart homes could be equipped with many appliances, such as security sensors, smart household appliances, which are connected to the Internet and could be remotely activated or disabled by a smartphone application, or even they are embedded with sensors that monitor whether a maintenance is needed. The smart speakers can provide the user a suitable interface with all these devices, switching on the air conditioner with voice command, or questioning about the state of a household appliance, whether a maintenance is needed, or an assurance is expired.

Hence the potential role of smart speakers in smart homes is defined, it is important for the manufacturers of these devices to study the market of the smart homes in general. The Internet of Things Observatory of Politecnico di Milano reports that in 2019 the Italian smart home market reached 530 million euros, increasing by 39% from the previous year (380 million euros), where the smart speakers accounts for 18% of the market (95 million euros), overcome only by the security solutions (28% of the market). Whereas, in Germany the market reached a value of 2.5 billion euros (+39% from 2018), 2.5 billion euros in United Kingdom (+47%) and 1.1 billion euros in France (+38%). The US market is still the most developed and its value is 84.6 billion dollars in 2020 (+15% from 2019), and it is expected to even growth with a growth rate (CAGR) of 18.2% between 2020 and 2023[17]. The figure below (Figure 4) shows the market increase of smart home devices and appliances for the European countries.

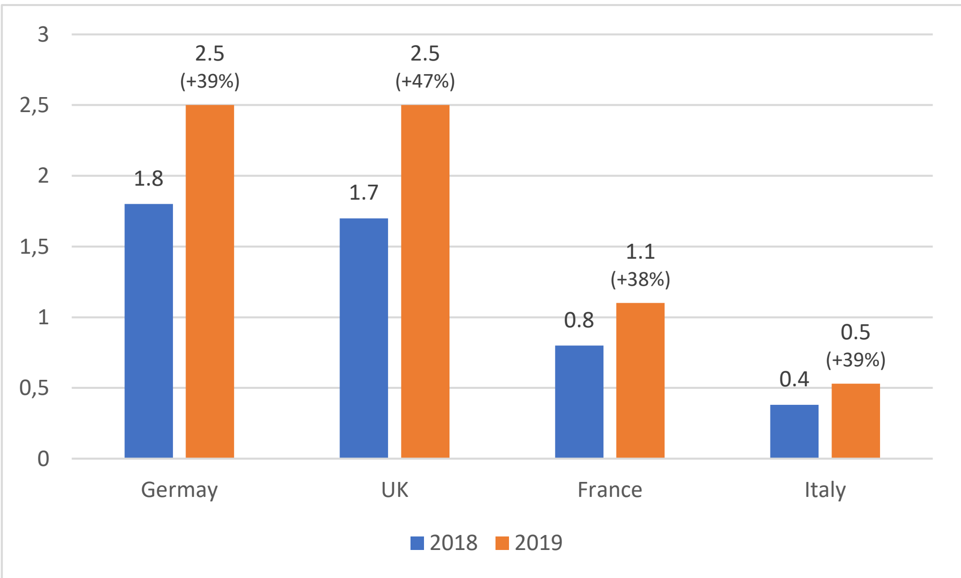

Figure 4: market value of smart homes devices and appliances in 2018 and 2019 in billion euros and the Year on Year (YoY) growth rate by European country. Source: our reworking of data by Internet of Things Observatory of Politecnico di Milano.

[17] Source: Smart-home worldwide, *Statista*. https://www.statista.com/outlook/283/100/smart-home/worldwide.

1.3. SMART SPEAKERS USAGE BEHAVIOURS

For the marketers, understanding the customers' usage behaviours is at least as important as the study of the market size. Bentley *et al.* (2018)[18] studied 65,499 interactions with Google Home devices and finds that the use is affected by the hour of the day. Use sharply increases from 6 to 7 am, having a second pick before lunchtime (between 12 and 1 pm) and a third one between 5 and 6 pm.

After the purchase, the users seem to go through two phases: *exploration* and a *consolidation*. The exploration phase lasts from the first two to three weeks, when the user tests what the smart speaker is able to do, what kind of questions could be queried and what needs the device is actually able to satisfy. Moreover, Sciuto *et al*. (2018) observed that "households attempted to test Alexa in different ways to see how intelligent the device actually was". During this period, the unique domains (e.g. music, information, home automation, and so forth) that the users query per day ranges from an average of 3 to 6. During consolidation phase, which follows the exploration one, the unique domain queries per day are stabilized around 3, meaning that the users are not more encourage on exploring the potentialities of their devices and they starts to use them for a certain task, e.g. to listen to the music, to set alarms, or to ask for news. At the same time, the queries become more specific, e.g. the users start to ask for specific features of the weather, such as the humidity or the high temperature of the day; likewise, they ask for specific songs and not to play music in general. It was even discovered that music is the domain more likely to be queried in average by the users belonging to all the ages, meaning that the smart speakers are still used as traditional speakers.

According to Voicebot Consumer Adoption Report (Jan. 2019), 38% of smart speaker owners use the device to listen to streaming music services in daily basis, 70% on weekly basis and 83% on monthly basis, making it the most common use case.

Sciuto *et al*. (2018)[19] revealed the same tendency of using the smart speakers mostly to listen to music (it accounts for 25.0% of total queries). Commands related to smart home domain, such as light control, accounts for 14.7%, while the weather information rest at 4.6%, while the purchases via smart speakers counts only the 0.3% of total queries. The Figure 5 exhibits the most common commands by users of smart speakers, from the research of Sciuto *et al*. About the physical placement of the devices, they noted that most of the participant of the study placed their smart speakers in their bedrooms and living rooms, followed by the kitchen and the home office.

[18] Bentley, F., LuVogt, C., Silverman, M. Wirasinghe, R., White, B., Lottridge, D. (2018).
[19] Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018, June).

Category	Keywords	Sample Commands	Count	
music	play, pause, song, skip, etc.	"play caskey on iheartradio" "alexa play jesus christ work Walmart" "alexa resume"	69,619	25.0%
other	—	"alexa start" "launch the bible app" "alexa turn that off"	56,446	20.3%
smarthome	office, kitchen, light, upstairs, etc.	"alexa turn the bedroom light off" "alexa turn on kitchen sink" "floor colors lights off"	40,870	14.7%
text not available	—	*Amazon was unable to transcribe due to misfire or low audio quality*	27,069	9.7%
weather	temperature, outlook, rain, etc.	"weather update" "what's the weather for tomorrow" "what's the weather in carlisle pennsylvania today"	12,731	4.6%
question	what's, where, why, where's	"alexa what's the difference between kosher salt and regular salt" "what is my drive to work look like"	10,409	3.7%
timer	timer	"set a timer for twenty minutes"	10,128	3.6%
wakeword	alexa, echo	"alexa" "echo"	9,715	3.5%
lists	shopping, list, add	"alexa add ritz crackers to the shopping list" "add girls shampoo to the grocery list"	7,714	2.8%
volume change	volume	"set the volume to five" "turn down the volume"	7,555	2.7%
personality	are you, is your, good morning, etc.	"good morning" "alexa good night" "alexa thank you" "simon says hi thank you so much"	7,366	2.6%
time	time	"what time is it"	6,219	2.2%
alarm	alarm	"alarm in fifteen minutes" "alexa set an alarm at nine thirty five a.m."	4,265	1.5%
news	news, briefing, flash, update	"stop the news" "ask cnbc if the latest news" "alexa daily briefing"	3,058	1.1%
joke	joke	"tell me a trump joke" "tell a joke" "tell me another joke"	2,819	1.0%
connectivity	bluetooth, pair, connect, sync	"alexa connect to phone" "connect to yamaha" "alexa yes pair"	1,788	0.6%
purchase	order, purchase, buy	"order cascade dash button" "order the dog complete series dvd" "alexa order scrabble twist"	883	0.3%

Figure 5: content categories for Alexa commands. Keywords were used to categorize the 278,654 commands. Source: Sciuto et al. *(2018).*

1.4. MARKETING VIA SMART SPEAKERS

The marketers can take advantage from the exponential growth of smart speaker market involving on three main activities: by voice advertisements (sometimes called "voice marketing", Hu *et al*., 2019), by developing their own application for smart speaker (the third-party skill) and by enabling the user to directly purchase offerings from their devices. In this scenario, the smart speakers represent a new touchpoint between the marketers and the customers, a one which could become very pervasive and persuasive.

Smith (2018) administered a questionnaire to young adults (age between 19 and 34) to analyse their attitudes toward several advertisement cues via smart speakers. She found that only 9% of the respondents do not want any marketing messages provided by their smart speakers, while the average number of messages that are considered acceptable in 1-hour period is around 2.6. "Provide information about availability and location of product/service" and "Tell me about a sale or discount" are the most appreciated messages, while the users are not interested on messages that "Contain my name" or are "Spoken by celebrities". About the functionalities of marketing messages, Smith found that the users really appreciate the possibility to skip the message or to tell about more detailed information, whereas they are less interested on saving the messages on the smart speakers or sharing them with others through social media.

The research even provides the users' willingness to hear marketing messages for specific product categories (Figure 6).

Product category	Often or Always	Seldom or Sometimes	Never
Travel	68%	23%	9%
Books, movies, & music	64%	31%	5%
Electronics	54%	40%	6%
Sports & outdoor products	52%	32%	16%
Groceries/food	51%	37%	12%
Clothing & shoes	47%	42%	11%
Healthcare (medical, exercise, diet)	47%	41%	12%
Home & kitchen	37%	49%	14%
Beauty & personal care	36%	42%	22%
Computers	32%	58%	10%
Pet products	28%	47%	25%
Automotive	24%	60%	16%
Apps & games	23%	57%	20%

Figure 6: Percentage of people willing to hear marketing messages for specific product categories on smart speakers. Source: Smith (2018).

Here, we can see that more than 60% of young adults appreciate marketing messages on their smart speakers when they talk about "travel" or "books, movies & music" and they are willing to hear these messages "often or always". The main implication of this research is that the users are willing to listen marketing messages on their smart speakers and the marketers can satisfy the users mainly by provide them useful information about promotions, sales and location of products/services.

Currently, Google do not allow advertisements and sponsored messages on its devices, due to the idea of building trust and iterating a great user experience. Likewise, Amazon rejects skills that "Includes or otherwise surfaces advertising or promotional messaging in skill responses, notifications, or reminders"[20]. However, some exceptions are provided by Amazon's policy for skills responses: streaming music, streaming radio, podcast, and flash briefing skills may include audio advertisements if the ad does not use Alexa's voice (or similar) and the content of the ad is not different from the same ad provided outside Alexa; skills that promote products/services to purchase, if the promoted offering is available in the skill platform itself; "skills may include audio messaging informing customers of promotional offers or deals in response to specific requests from customers"; "skills that are specifically designed to promote a product or service", to strictly advertise for those product/services.

The idea behind the change of voice when advertise for an offering is that the user should be able to understand that the message concerns an advertisement, and not a common interaction with the voice assistant, resulting in increasing trust toward the device. Moreover, Kim *et al.* (2018) even found that voice dissimilarity between the voice assistant and the content announcer can enhance the

[20] Alexa Skills Kit. *developer.amazon.com*. Policy Testing for an Alexa Skill, Advertising (5), from https://developer.amazon.com/it-IT/docs/alexa/custom-skills/policy-testing-for-an-alexa-skill.html#5-advertising.

effectiveness of the ad in terms of memory and retrieval accuracy, while the contextual relevance significantly moderates the relationship between the voice similarity and the memory accuracy.

Thus, nowadays the greatest possibility for marketers is represented by the development of their own third-party skill, which enables the firm to promote its product by voice advertisements and to sell products/services directly from the smart speaker. These skills could work as shopping platforms with which the user could interact by voice command to make purchases. Many online stores have already built their skill and make that available for the users, namely BestBuy[21], which sells electronic and technologic products and home furniture; Fandango created a skill to discover new movies, browse movie times and buy tickets[22]; OpenTable made its restaurant reservation service in U.S. available on smart speakers[23]; and more other activities could shift their efforts from traditional and online touchpoints to the smart speakers, developing their third-party skill.

1.5. THEORETICAL AND PRACTICAL CONTRIBUTIONS

Once the smart speakers have been recognized by many marketers as a new and fast developing marketing channel, studying users' behaviours and adoption mechanisms becomes crucial to improve their business and extract more value from both potential and current customers. Therefore, many studies concerning users' acceptance of smart speakers were conducted in the last years.

Kowalczuk, (2018) administered a survey questionnaire to 293 German and found that the intention to use the smart speakers is strongly affected by the perceived enjoyment and the perceived usefulness of the devices, while the perceived ease of use has not a significant effect on the acceptance.

Chu (2019) confirmed the importance of enjoyment and perceived usefulness on the adoption of smart speakers, even discovering a significant effect of the perceived ease of use on both the other two variables. Other important antecedents of acceptance that emerge from those studies are the trust and the risk perception (on privacy and security).

Sohn & Kwon (2020) studied the acceptance of IoT devices and appliances in order to find the acceptance model that better explains the intention to use. They found that the TAM (proposed by Davis, 1989) is not the best model to explain acceptance since the behaviour intention is explained

[21] From: https://www.bestbuy.com/site/misc/using-voice-assistants-to-explore-best-buy/pcmcat1527688377766.c?id.

[22] From: https://www.fandango.com/canvas/alexa-faq.

[23] From: https://blog.opentable.com/2017/opentables-updated-alexa-skill-is-now-available-on-amazons-echo-show/.

only by perceived ease of use and perceived usefulness. Whereas the VAN[24] is the most effective analysed model to explain the AI-based product acceptance.

Since many efforts to understand the users' acceptance of smart speakers have already been spent, this work is focus on the adoption mechanisms and antecedents of one of the major opportunities provided by these devices, I mean, the recommendations that the devices can offer to the users. As we will see (chapter 2), recommendations are an effective tool to enhance user experience and cope with the information overload, which affects users' decision making on most of the online environments. Recommendations could be provided by smart speakers through software tools called Recommender Systems (RS).

Many studies deal with the issue of measuring the efficacy of the recommendations (paragraph 2.3); whereas many others concern the acceptance of recommender systems in several situations. Wang *et al.* (2012)[25] employed the UTAUT (Unified Theory of Acceptance and Use of Technology) acceptance model, developed by Venkatesh *et. al.* (2003)[26], in order to study the variables that affect users' behavioural intention to use different types of RSs (collaborative filtering and content-based). They found that "PE [Performance Expectancy] and Trust are two major concerns for those who use the content-based system. On the other hand, SI [Social Influence] and Trust are another two major concerns for those who use the collaborative filtering recommender system".

Armentano *et al.* (2014)[27] studied the acceptance of context aware recommender systems building a model based on Technology Acceptance Model (TAM). They found that "although the system interface was the same for all users, users found the context aware system easier to use, with more attractive recommendations that were better adapted to their mood and tastes", with respect to a generic collaborative filtering RS.

Pu *et al.* (2011)[28] attempted to build a user centric evaluation framework for users' acceptance of recommender systems and their intention to purchase the recommended items, and they called the model ResQue. This model is analysed in paragraph 2.3.3.

The quoted researches attempt to investigate the user's acceptance of recommender systems or the acceptance of smart speakers separately one each other. However, to our knowledge, no studies concerning users' acceptance of recommender systems from smart speakers are been conducted, although this theme has crucial implications for the marketers who want to exploit the advantages provided by RSs on their third-party skills.

---

[24] Kim, H. W., Chan, H. C., Gupta, S. (2007).
[25] Wang, Y. Y., Townsend, A., Luse, A., Mennecke, B. (2012).
[26] Venkatesh, V., Morris, M. G., Davis, G., & Davis, F. D. (2003).
[27] Armentano, M. G., Abalde, R., Schiaffino, S., Amandi, A. (2014, November).
[28] Pu, P., Chen, L., & Hu, R. (2011).

This study aims to fill up this gap and to provide a comprehensive insight of users' acceptance of such systems from smart speakers, underlying the main variables that affect the acceptance and the relationships among them. The deeper understanding of the variables that affect the acceptance of recommender systems from smart speakers and the acceptance model proposed on this study represent the main theoretical contributions of this work, since no previous researches involved the study of acceptance of RSs from smart speakers. Then, practical contributions are provided with the suggestions, inferred by the results of the data analysis, for the marketers who want to improve their business developing recommendation systems for smart speakers.

1.6. WORK OVERVIEW

The research in structured as follows. In chapter 2 an introduction to recommender systems is provided, together to how they work, what are the main methodologies applied in such systems and their potential advantages both for customers and marketers. In paragraph 2.1, the content-based algorithms are explained, while content filtering ones are dealt with in paragraph 2.2, where item-item and user-user approaches are explained. The different approaches that are used to evaluate the efficacy of recommender systems are provided in paragraph 2.3. Here, the accuracy is one of the first metrics ever studied to evaluate the performances of RSs (2.3.1), while in the last years more metrics become object of interests (2.3.2), such as the coverage, the serendipity and the diversity. The last step for a comprehensive evaluation of RS was made with the introduction of a user-centric evaluation framework (paragraph 2.4), which introduces the user perspective and intention to use the RS.

The chapter 3 represents the core contribution of this research and deals with the users' acceptance of recommender systems from smart speakers. To answer the research questions, a theoretical framework was built based on previous research studies on the field of technology acceptance (paragraph 3.2), taking the Davis's Technology Acceptance Model as the reference point for the developing of a more complex guesswork. The variables of the model are identified and structured in a survey questionnaire that was administered to 286 respondents (paragraph 3.3). Once the data was gathered, a Structural Equation Modeling (SEM) analysis was performed in R program in order to answer the research questions (paragraph 3.4). It was found that the Perceived Usefulness and the Enjoyment are the strongest drivers of acceptance of recommender systems from smart speakers, while the Perceived Ease of Use, one of the former variables of the TAM, seems not to have any significant effect on the other variables.

In chapter 4, some suggestions for the marketers are provided, based on the results of the research (paragraph 4.1). The issue of the users' trust toward RS from smart speakers is more deeply investigated (paragraph 4.2). Then, the limitations of the research are disclosed, while suggestions for future researches are provided (paragraph 4.3).

2. RECOMMENDER SYSTEMS

Recommender systems (RS) are software tools and techniques which aim to provide to the user with suggestions for items based on their expected preferences. According to the technique, the expected preferences are computed in several ways by the algorithms. The suggestions that are provided to the user are a small subset of the whole set of available items, resulting on a significantly decrease of mental effort of the search. The “items” should be anything the user is looking for, such as a movie to see, a music to listen to, a restaurant where having dinner with her/his partner, a hotel where have holidays, or whatever product or service is offered by the provider of the system.

In the last decades, the interest on RS sharply rises, since more consumers’ data could be stored and analysed. Nowadays, recommender systems are embedded in most of the e-commerce websites, such Amazon.com, or Netflix, which operates in streaming video industry.

The interest is based on the several advantages provided by recommender systems. From user perspective, using a recommender system results in easier filtering and navigation through the website, a decreasing information overload and, thus, overall decrease of mental effort. Personalized contents could sharply increase customer satisfaction, together to the delight and, in second instance, the loyalty toward the RS provider.

From the firm standpoint, the automatic operations of RSs can increase profits without heightening the marketing costs, resulting in higher ROI (Returns on Investments). Higher customer satisfaction and loyalty, then, result in a more robust customer base, which is a strong competitive advantage, since the clients are less likely to switch to competitors’ offerings. The RSs are even a good way to cope with the so-called *long tail phenomenon*. It is well-known that offline stores suffer the limitations due to their storage spaces, while the online ones could expose the potential clients to an almost infinite products range, offering the opportunity to buy also the less popular products. Therefore, RSs could benefit from long tail phenomenon, which explains that most profits come from less popular product (the *long tail*). The Figure 7[29] shows the relationship between the sold products and their popularity.

[29] Source: Marketing Analytics. *DataMiningApps*. From: http://www.dataminingapps.com/ dma_research/ marketing- analytics/.

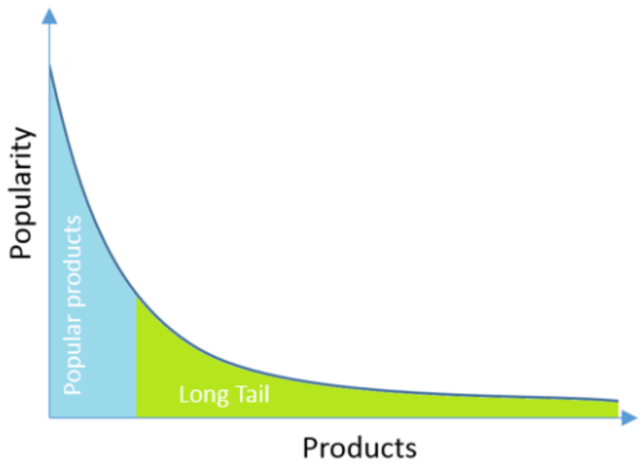

Figure 7: In blue, the most popular products accounts for a small percentage of all the other products (the long tail in green).

To understand how the RSs are important for the development of the businesses, McKinsey states that 35% of purchases on Amazon are the result of their RS; 70% of the time the people spend on watching videos on YouTube refers to its RS, while 75% of the movies watched on Netflix come from its RS[30]; the Streaming Video On Demand (SVOD) algorithm helps Netflix to save $1 billion each year[31].

The techniques employed by the algorithms to compute users' expected preferences could be grouped in 6 macro-approaches:

1. **Content-based**: the expected preferences are inferred from the attributed of items that have been already chosen by the user in past interactions.
2. **Collaborative filtering**: the expected preferences are inferred from the items that have been chosen by users with similar characteristics.
3. **Community-based**: suggestions are based on the preferences of user's friends.
4. **Demographic**: this type of recommendations is based on users' demographic profile.
5. **Knowledge-based**: these systems recommend items based on specific domain knowledge about how certain items features meet user needs and preferences and, ultimately, how the item is useful for the user[32].

[30] Rodrìguez, G. (2019).
[31] Subramanian, V. (2019).
[32] Ricci, F., Rokach, L, Shapira, B., Kantor, P. B. (2011).

6. **Utility-based**: this type of system makes recommendations based on a computation of its usefulness for each individual user[33]. This approach implies the calculation of user utility curve to infer his/her preferences.

Moreover, **hybrid** RS are combinations of two of the above techniques. These are usually employed to take advantage of the pros of different approaches and, meanwhile, to fix the disadvantages of each one.

The archetypes of recommender systems are the content-filtering and the collaborative-filtering. Therefore, we will deeply analyse them on the following paragraphs.

2.1. CONTENT-BASED RS

The content-based approach refers to all those algorithms which infer the user's expected preference from the characteristics of the items the user have already chosen in the past. The assumption, here, is that products that are similar to others the user have already chosen are more likely to be preferred than items which differ from them. The user preferences have to be consistent along the time, since the algorithm assumes that past preferences are good predictors of future ones.

The Figure 8[34] displays how content-based recommendation systems work.

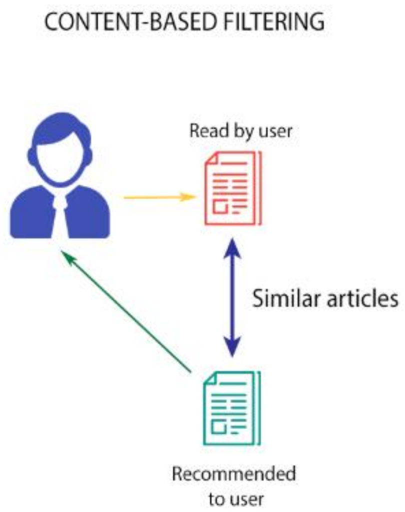

Figure 8: Content-based algorithm relays on users' profiles and past preferences to recommend items similar to the ones already chosen by the user.

[33] Underwood, C. (2020).
[34] Source: Doshi, S. (2019).

Items are rated by a content analyser along several features, which generates an item vector. Likewise, a user profile is created relying on previous behaviours and choices. Then, the similarity between the user vector and the item vector could be computed. Although several measures and indices have been created, the cosine similarity index is the most employed on this type of computations. The cosine index formula is the following:

$$\cos(\theta) = \frac{\sum_{i=1}^{m} x_i y_i}{\sqrt{\sum_{i=1}^{m} x_i}\sqrt{\sum_{i=1}^{m} y_i}}$$

Here, x is the user vector and y is the item vector. The cosine similarity metric is computed along all the m features, then the obtained ratings are sorted in decreasing order and only the top k items compares on the recommended list. In opposition to the Euclidean distance, the cosine metric doesn't account for the magnitude of the features employed on the computation, therefore it is used to predict the similarity of texts of different size (number of words). The importance of text classification and analysis is fundamental for content-based algorithms since most of the items descriptions on the Internet are made up textual contents.

To provide an example of how a content-based RS works, we can consider a book offer. The books could be classified into several genres (the features), then a binary value could be attributed to each feature according to the genre of the book. If the book belongs to the genre, it scores 1, otherwise it scores 0. A $n \times m$ matrix is created, where the rows contain the books and the columns contain the genres. Likewise, the user vector is computed according to the genres of the books he/she has already read. Each book vector from the recommendation matrix is compared to the user vector employing the cosine metric formula, then they are sorted in decreasing order. The top k books are the ones most similar to the user's profile, therefore they will be recommended.

An alternative approach to analyse textual similarities is to consider a document-term matrix, where each row represents the item (in our example, a book) and the columns represent a word. the matrix is filled by the number of times a word compares in each book. Intuitively, books that use similar set of words belong to the same genre, then it is possible to attribute similarities between user's profile and item vectors according to the number of times a set of words appears. However, this count must be weighted by the importance of each word. To do so, it is usually employed the TF-IDF function. The assumption, here, is that the words that compare many times in the document (Term Frequency, TF), but not in the others (Inverse Document Frequency, IDF), are more likely to identify its genre. The user preference for a book could be derived from the cosine similarity between his/her profile and the document-term matrix computed with TF-IDF.

The content-based approach is endowed with several pros and cons that has to be considered when it is employed.

One of the advantages of this approach is that it does not rely on community ratings, meaning that the introduction of a new item is comfortable, since only the item features have to be provided. On the other hand, the introduction of a new user could be problematic, due to the fact that user's behaviours are unknown. A second advantage of these systems is the transparency of the recommendations. It is suitable to explain why the recommended item was provided, simply referring to user past interactions ("The item was offered, since it is similar to products you have already purchase!"). Contrariwise, the collaborative filtering has the characteristic of a black-box, where it is difficult to assess a reason why a certain item is recommended (it is only because other people similar to the user prefer the suggested item, in combination to items the user prefer).

However, content-based RS suffers some limitations, such as a domain knowledge is usually needed. A sufficient number of features should be considered to provide an effective recommendation and no crucial variables should be omitted. Another issue related to these types of algorithms is that they suffer the over-specialization. Since they rely only on items similar to the ones the user has already exhibited a favourable behaviour, it is pretty unlikely that they suggest items that differ from those ones, though the users are interested to. To explain this point, we can consider the example of a new user of a streaming video platform which employs content-based RS to suggest movies to their clients. If this user chooses to watch a western movie, the RS will recommend to the user only other western movies, even if the user likes also thrillers and horror movies. This is an over-simplification of the issue, but it explains why the content-based RSs suffer the specialization, and they do not help the user to explore and to look for new interesting items.

2.2. COLLABORATIVE FILTERING

Collaborative filtering approach refers to all those algorithms which infer the user's expected preference from the preferences of similar users. The assumption, here, is that if two users exhibited similar tastes in the past, it is pretty likely that they will share the same trend even in the near future behaviours (e.g. in the next purchase).

The Figure 9 below displays how collaborative filtering algorithms work.

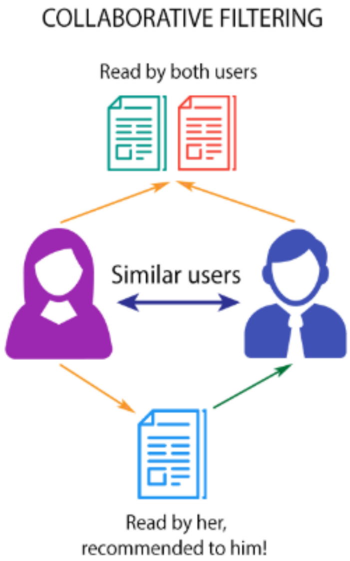

Figure 9: Collaborative filtering algorithm relays on users' past behaviours similarities to recommend items.

If book i was read by both user A and user B, and user A read also the book h, it is likely that the user B is interested on reading book h too.

The similarity measures could be several. According to the specific case, it could be employed the Pearson correlation to assess the "distance" between users or items. It is computed as follows:

$$s(a,u) = \frac{\sum_{i \in I_a \cap I_u} (v_{ai} - \bar{v}_a)(v_{ui} - \bar{v}_u)}{\sqrt{\sum_{i \in I_a \cap I_u} (v_{ai} - \bar{v}_a)^2 \sum_{i \in I_a \cap I_u} (v_{ui} - \bar{v}_u)^2}}$$

$s(a,u)$ is the Pearson coefficient between user a and user u; I_a are the set of items rated by user a, while I_u are the set of items rated by user u; v_{ai} is the rating that was assigned by user a to item i, while $\bar{v}_a$ is the average rating assigned by user a.

Shardanand (1994)[35] suggested to compute the similarity between users by the mean squared difference based on the mean difference of the items that both have rated. The formula was the following:

$$msd(a,u) = \frac{\sum_{i \in I_a \cap I_u} (v_{ai} - v_{ui})^2}{|I_a \cap I_u|}$$

[35] Shardanand, U. (1994).

Where $|I_a \cap I_u|$ is the number of items rated both by user a and user u.

Even the cosine similarity could be employed (that we have already faced in paragraph 2.1).

The collaborative filtering techniques are usually distinguished in two branches: user-user and item-item collaborative filtering. In the following paragraphs we describe the two approaches.

2.2.1. User-user collaborative filtering

According to user-user collaborative filtering, an item is recommended to user A if it was positively rated by user B, who exhibited taste similar to the user A's one. These algorithms follow three steps: (1) similarity between the users are computed; (2) a subset of users (the "neighbours") is selected according to users' similarities; (3) The prediction is computed according to neighbours' ratings. Therefore, the definition of similar users is fundamental to employ this approach. Similar users could be grouped using different types of algorithms, such as K-Nearest Neighbours (KNN), Support Vector Machines (SVM), artificial neural networks and decision trees (CARTs).

Here, an issue arises. To be performed, many algorithms need the definition of a distance between users in a n-dimensional space, where n is the number of items. The issue is that only few items are rated by the users, therefore the user-item matrix, which is employed to compute distances between users, in most cases suffers the problem of the *sparsity*. To cope with this issue, before computing the distances some pre-processing techniques are applied, such as Principal Component Analysis (PCA) or Singular Value Decomposition (SVD) in order to decrease the dimensionality of the user-item matrix and, thus, to prevent the problem of the sparsity.

Once the similar users have been defined, the items which had been positively rated by them are suggested to the reference user. The user's expected preference (or rating) could be calculated as a weighted mean of the KNNs, where the weights are represented by the distance between the reference user and his/her neighbour.

2.2.2. Item-item collaborative filtering

Instead of focusing on similar users (as the user-user techniques do), it is possible to infer the user's expected preference by looking at the chosen items. If other users usually chose the item 1 in combination to item 2 (not necessarily at the same time) and user A have chosen item 1, it is logical to suppose that user A may have a positive attitude toward item 2 too, thus item 2 is suggested.

Item-item collaborative filtering differs from content-based method, since the similarities between the items are not based on characteristics addressed *ex ante*, but they are the result of a community behaviour toward the offering. Here, as in the other cases, the behaviour could be both an explicit and an implicit user's preference.

User-user and item-item methods address different aspects of the data, then a combined approach could result in an increase of the overall accuracy of the prediction. Wang *et al.* (2006)[36] attempted to create a unified approach where the ratings are estimated by fusing predictions from three sources: predictions based on ratings of the same item by other users; predictions based on different item ratings made by the same user; ratings predicted based on data from other but similar users rating other but similar items. They demonstrated that a comprehensive approach leads to more robust predictions against the sparsity problem and more accurate results than the other methods.

Same results, but with a different combination of item-item and user-user filtering techniques, was reached by Ricci *et al.* (2011)[37], where the two approaches are optimized simultaneously (and not separately as in previous studies). They tested the model on Netflix Prize competition dataset, showing that the accuracy is indeed better than the other two approaches when they are employed separately.

2.2.3. Advantages and disadvantages of collaborative filtering

The main advantages of collaborative filtering techniques, with respect to content-based ones, are that they do not require domain knowledge, since the predictions are inferred from other users' behaviours and not on previously defined characteristics of the items. Another advantage is that this kind of algorithms offer the user a broader range of items, not relying only on the characteristics of the items have been already chosen. This mean that the user could benefit from the differentiation of the offering and not just repeating the same genre of books to read or movies to watch, which could be boring and annoying from the user's standpoint.

On the other hand, the limitations of content filtering are mainly related to the necessity of ratings (explicit preferences) or other kind of implicit preference. Since the data suffers the sparsity issue, a content filtering technique (in particular the memory-based ones, in opposition to model-based approaches) require many data both for items and users. Moreover, when a new item was provided, it has not any rating from which the preferences could be inferred. This last is not an issue for content-based algorithms, since the characteristics of the items are addressed *ex ante*.

[36] Wang, J., De Vries, A. P., & Reinders, M. J. (2006, August).
[37] Ricci, F., Rokach, L, Shapira, B., Kantor, P. B. (2011).

2.3. EFFICACY OF THE RECOMMENDATIONS

2.3.1. The accuracy

According to the recommender task, the efficacy of the recommendations could be evaluated along different dimensions. The accuracy metrics are the first studied measures of performance of recommender systems. Here, Herlocker *et al*. (2004)[38] stated that "an accuracy metric empirically measures how close a recommender system's predicted ranking of items for a user differs from the user's true ranking of preference". They distinguish three sets of accuracy metrics for collaborative filtering recommender systems according to the type of the ratings: (1) predictive accuracy metrics; (2) classification metrics; (3) rank accuracy metrics.

The *predictive accuracy metrics* "measure how close the recommender system's predicted ratings are to the true user ratings". These are good on evaluating users' ratings when they are evaluated in a continuous scale, or in a scale which is made up several possible ratings, such as 5-point scale, that is usually employed on quantifies the preferences (e.g. stars scales from 1 star to 5 stars). The most important of such indexes is the Mean Absolute Error (MAE), computed by the following formula:

$$MAE = \frac{\sum_{i=1}^{N} |p_i - r_i|}{N}$$

Where N is the number of relevant rankings and $|p_i - r_i|$ is the absolute difference between the predicted rank and the actual one. This is an easy measure to compute and interpret, however has the disadvantage that does not penalize huge differences between the prediction and the actual ranking. To accomplish this task, Mean Squared Error (MSE) could be employed, computed as it follows:

$$MSE = \frac{\sum_{i=1}^{N} (p_i - r_i)^2}{N}$$

The *classification accuracy metrics* "measure the frequency with which a recommender system makes correct or incorrect decisions about whether an item is good". These metrics are good predictors of accuracy when the preferences are dichotomous (buy/not-buy, watch/not-watch, read/not-read). Here, to measure the accuracy of the system, a classification matrix (even called "confusion matrix") is created to compare predictions and actual values, as in Table 2.

[38] Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004).

	Recommended	Not-recommended
Selected	TP	FN
Not-selected	FP	TN

Table 2: classification matrix. This matrix is used to compute classification accuracy metrics.

True Positives (TP) are the number of times the recommendation matches the users' actual preferences and the users exhibit a positive behaviour toward the item. True Negatives (TN) are the number of times the recommendation matches the users' actual preferences and the users exhibit a negative behaviour toward the item. False Positives (FP) are the number of times the recommended item is not in the users' preference set. False Negatives (FN) is the number of times the item is not recommended, though the user exhibits a preference toward the item.

Here, two main classification accuracy metrics are usually employed: Precision and Recall. The precision (P) of the recommender system is the ratio between the true positives and the total number of items recommended ($TP + FP$), thus it represents the probability that a recommended item is selected.

$$P = \frac{TP}{(TP + FP)}$$

The recall (R) is the ratio between the true positives and the total number of items selected ($TP + FN$), Therefore, it represents the percentage of well-classified users' preferences (also known as *true positive rate*).

$$R = \frac{TP}{(TP + FN)}$$

The higher Precision and Recall are, more effective is the recommendation system on matching users' preferences.

The *rank accuracy metrics* "measure the ability of a recommendation algorithm to produce a recommended ordering of items that matches how the user would have ordered the same items" (Herlocker *et al*., 2004). These metrics differ from classification accuracy metrics by the fact that the rank is not a binary response. It is a prerogative of all recommender systems to offer the users "the best recommendation possible", therefore a rank accuracy metrics need to assign a higher weight to the first item of the recommendation list.

2.3.2. Other efficacy measures

While in the first generation of recommender systems the efficacy was evaluated only on terms of prediction accuracy, in the last years other measures become popular and aim to account for different recommendations tasks. Ziegler *et al.* (2005)[39] approached the issue, stating that not only the accuracy of the recommendation list is a predictor of its quality, but even the *diversification* of the items that offers is a proper measure of performance. Here, the diversification "refers to all kind of features, e.g., genre, author and other discerning characteristic" of the offering. The possibility for the user to freely select the item that best fits his/her interests is one of the tasks a recommender system needs to accomplish, otherwise the system could be perceived as a censor agent. We have already observed that the issue of the topic diversification is one of the limitations of content-based recommender systems, while the collaborative filtering ones are more appropriated to cope with this problem.

Ziegler *et al.* introduced also the Intra-List Similarity (ILS) metric to assess the topic diversification of the recommendation list. It is computed as follows:

$$ILS(P_{w_i}) = \frac{\sum_{b_k \in P_{w_i}} \sum_{b_e \in P_{w_i}, b_k \neq b_e} c_0(b_k, b_e)}{2}$$

Where P_{w_i} identifies the recommendation list, b_k and b_e are two different items belongs to the list and $c_0(b_k, b_e)$ is an arbitrary function measuring the similarity (with values equal to -1 for dissimilarity and equal to +1 for similarity) between two items, according to some custom-defined criterion.

Another metric used to evaluate the efficacy of the recommender systems is the so-called *serendipity*. According to McNee *et al.* (2006)[40], the serendipity of a RS is "the experience of receiving an unexpected and fortuitous item recommendation". It is sometimes referred as a *novelty* measure, however, the serendipity mainly concerns the emotional reactions to the recommendation, such as surprise and astonishment. Since popular items are more likely to be preferred, the novelty and serendipity measures attempt to avoid banal recommendations, adjusting the expected preference for the popularity of the item.

[39] Ziegler, C. N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005, May).

[40] McNee, S. M., Riedl, J., & Konstan, J. A. (2006, April).

Another metric associated to the efficacy of RS is the *coverage*. This measure is used in the literature to refer to two different concepts, as Ge *et al.* (2010)[41] mentioned: (1) the percentage of the items for witch the system is able to generate a recommendation (also known as *prediction coverage*); (2) the percentage of the available items that actually are ever recommended to the users (also known as *catalogue coverage*). In order to take the more advantages as possible from the long tail phenomenon, it is more profitable for the provider of the system to have a high coverage (in both the meanings quoted above).

The prediction coverage (PC) could be formulated as it follows:

$$PC = \frac{|I_p|}{|I|}$$

I denotes the set of available items, while I_p is the set of items for which a prediction could be made.

Whereas, the catalogue coverage (CC) is computed with the following equation:

$$CC = \frac{|\bigcup_{j=1}^{N} I_L^j|}{|I|}$$

Here, according to Ge *et al.* (2010), I_L^j denotes the set of items contained in the list L returned by the j^{th} recommendation.

2.3.3. Toward a user-centric evaluation framework

Till now, we have not explicitly considered the user experience to evaluate the RSs performances. The interest on such an issue came more recently and in the last years come user-centric evaluation frameworks for RSs were developed. Pu *et al.* (2011)[42] suggest a model called *ResQue* (REcommender Systems' Quality of User Experience), which has the objective of "measuring the qualities of the recommended items, the system's usability, usefulness, interface and interaction qualities, users' satisfaction with the systems, and the interface of these qualities on users' behavioural intentions, including their intention to purchase". Their research is based on the development of previous models, developed to evaluate the usability of information systems, such as

[41] Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010, September).
[42] Pu, P., Chen, L., & Hu, R. (2011).

the TAM (Technology Acceptance Model) by Davis (1989)[43] and the SUMI (Software Usability Measurement Inventory) by Kirakowski (1993)[44]. The proposed model is divided in four blocks, where the preceding blocks affect the following ones, which are: (1) users' perceived quality of the system; (2) users' beliefs toward the system; (3) users' attitude; (4) behavioural intentions.

The user perceived quality refers to users' perceptions of objective characteristics of the RS, which are the perceived accuracy of the recommendations, the novelty and the possibility to discover new interesting items, the attractiveness (which refers to the capability of recommended items to stimulate users' imagination), the diversity of the items that compare in recommendation list and the context capability (whether or not the recommended items fit the general or personal context requirements).

The users' beliefs refer to the transparency of the systems (e.g. the explanation of why the items are or not recommended to the user), the sense of control, the perceived usefulness and de perceived ease of use of the system.

Users' attitudes comprehend the trust and confidence toward the system and the overall user satisfaction; while the behavioural intention was studied both for purchase and the use intention.

The Figure 10 shows the ResQue model and how the variables should be studied in combination to have a comprehensive evaluation of the recommender systems from a user perspective.

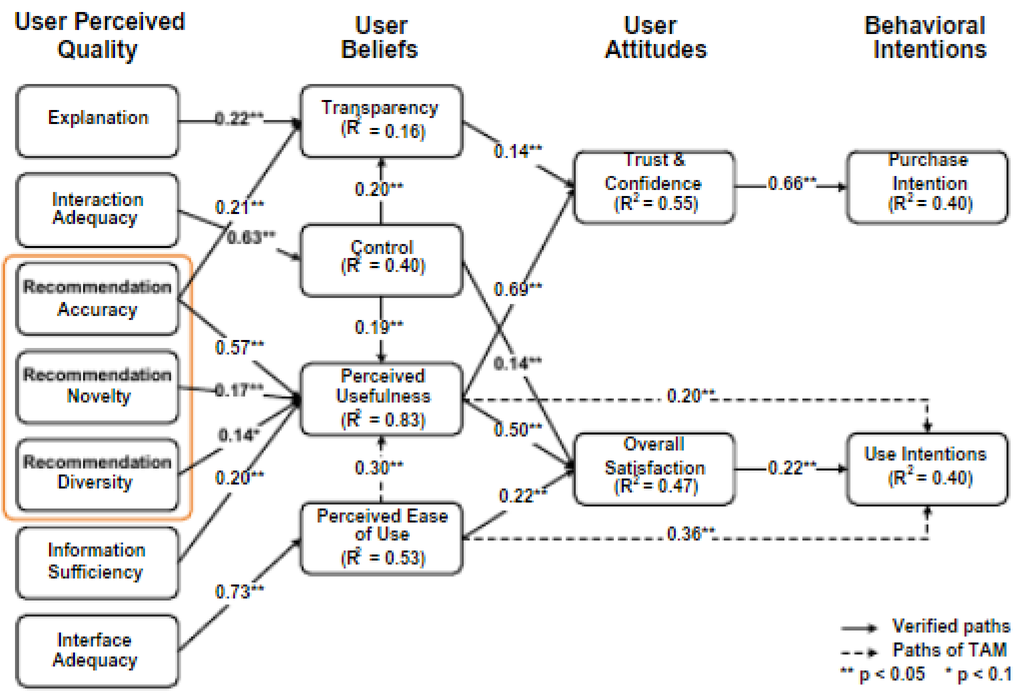

Figure 10: the ResQue model by Pu et al. (2011). The behavioural intention is the result of a four-steps evaluation which starts from user's perceived quality of the system, to go through the user's beliefs and attitudes. In orange box, the measures of effectiveness of RSs usually employed are highlighted.

[43] Davis, F. D. (1989).

[44] Kirakowski, J., & Corbett, M. (1993).

The study by Pu *et al*. underlines the necessity to evaluate the quality of recommendation lists along many dimensions that ranges from the layout of the web portal that is equipped by the system, to the user's overall satisfaction toward the items which compares in the research. The whole interaction of such variables results on the acceptance of the system and its usage, or in purchase intention. The users' "conversion" is one of the objectives of the RSs from the firm perspective, since it means that the software is able to provide new consumers and retains the current ones. Here, the "conversion" is the action made by the user that results from marketing stimuli and it is the objective of the recommendations (e.g. the user purchase or select an item suggested by the RS).

2.4. PECULIATITIES OF RS FOR SMART SPEKERS

The recommendation list can be provided in different ways. It is possible to display only the top *k* items of the entire set, according to a similarity metric, or only the most preferred one. Otherwise, a threshold approach could be employed; here, only the items which overcome a certain threshold of the similarity metrics will be actually recommended. One of the issues concerning recommendations from smart speaker is that such devices have not any screen in which the recommendation list could be displayed. Since smart speakers use only the voice to communicate to the user, they are forced to recommend only the first item of the list, namely, the one which is more likely the user is interested on. Because we have already explained that the diversity and the coverage of the recommendation list is a source of value both for customers and the firms, the marketers need to cope with this severe limitation. One possibility is to equip the smart speakers with more flexible conversational capabilities. A study conducted by Jung *et al.* (2019)[45] suggests that "people found it useful to get answers on their questions by talking with the assistant", and this capability could be exploited in order to equip the smart speakers with conversational recommender systems. Image to look for a restaurant to have dinner and ask to your smart speaker for a hint: "Alexa! Where can I have dinner tonight?" Since we image that the smart speaker is equipped with a recommender system and it has a well-established user's profile, it knows that you use to go to the pizzeria on Saturday night with your partner, then it suggests: "You can go to Aldo's Pizza tonight. It is the best pizzeria in the neighbourhood!" However, today you are not interested on eating pizza, you prefer something different, then you attempt to ask: "No, today I'm looking for something else. Can you suggest another restaurant?" Here, a problem arises. The recommender systems would have offered you a

[45] Jung, H., Oh, C., Hwang, G., Oh, C. Y., Lee, J., & Suh, B. (2019, May).

pizzeria, but there is a conflict with your previous statement which impose to find a restaurant that *is not* a pizzeria. Then, the recommender systems need to understand that your question is contextualize and the eligible recommended items should be filtered by "*restaurant-such-as-it-is-not-a-pizzeria*". This is a task difficult to accomplish, and nowadays the smart speakers are able to interpret only single commands and not sequential ones.

The possibility to be involved on more complex conversations, as the human being usually do when speaking one each other, can simulate a recommendation list wider than a one-item list that results from a single command and relative answer, which is the paradigm of current smart speakers. Even preference revision is a task that could be implemented with more flexible conversational mechanisms. Indeed, Smith (2018)[46] observed that smart speakers users appreciate the possibility to interact with promotions via smart speakers in three different ways: repeating and skipping the marketing message; asking for more information about the promoted item.

Even the explanation of why a certain item was suggested and how the system works increase the users' trust and confidence, then it results on higher quality of recommendations. Tintarev & Masthoff (2007)[47] demonstrated that the ways the explanation interface affects users' attitude toward the system and its persuasiveness. The problem, here, is that the smart speakers has not a user interface (a screen) out of the voice. Further researches should be conducted to discover the best way to explain by voice the reasons of a recommendation, avoiding annoying the user, or whether this implementation is useful at all for enhancing user experience in the case of smart speakers.

Another issue concerns the physical placement of smart speakers. Sciuto *et al.* (2018)[48] correctly note that "user must have a felt sense that the conversational agent is available in a particular space", a feeling caused by the fact that these devices lack a screen and their physical footprint is small. For these reasons, the environment where the users place their smart speakers affect the use. Thus, it is crucial for recommender systems from smart speakers to introduce some sort of context-aware technique. Context-Aware RSs (also called CARS) provide suggestions based on several contextual information, such as the date and hour, the location, the season, but even the user's mood and current needs could shape the context. They are usually grouped in three main approaches: (1) contextual pre-filtering; (2) contextual post-filtering; (3) contextual modelling.

In *contextual pre-filtering*, the offering inventory is filtered by the items that fit the context, then the recommender algorithm runs only on eligible items to create the recommendation list. In *context post-filtering*, the recommendation list is produced at the first step, then only the items that fit the

[46] Smith, K. T. (2018).
[47] Tintarev, N., & Masthoff, J. (2007, April).
[48] Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018, June).

context are actually suggested to the user. In *contextual modelling*, the context compares as a variable in the recommender model, therefore it is encapsulated on the calculations of similarities between users or items.

According to the room where the smart speaker is placed, it seems logical to suppose that it is expected to accomplish different tasks, then the recommendation algorithms could take advantage from this information to provide more accurate suggestions. This information could be partially used even to cope with the so-called *cold start* issue, which affects the RS. The cold start refers to the poor user's information at the first stages of the usage of the system. New users lack a consolidated profile, therefore it is not possible to provide him/her with personalized recommendations, until the number of user-system interactions rises.

An important context information is even the hour of the day, since it was discovered that the usage behaviours significantly change from morning to the night[49].

Perhaps, even the voice tone could be considered a context information. From the voice tone, indeed, the mood of the user can be inferred, and this information should be exploited to provide him/her with more accurate recommendations. In fact, it was observed that the people's mood strongly affects their decision making[50]. Voice tone could be also used to count the number of people in the room and who they are. This contextual information could be used to maximize the effectiveness of the recommendation for a group of people, and not only for the individual, computing the prediction of the expected preference at the group level.

The lack of a screen is even more a limitation when the user needs to see a product, perhaps to purchase it. Voice interactions are sometimes not enough to explain characteristics of items, or it is not the most effective method; in these circumstances, a visual representation becomes indispensable. To face this problem, the smart speaker should be connected to a screen, perhaps to the user's smartphone or to a smart-TV.

[49] Bentley, F., LuVogt, C., Silverman, M. Wirasinghe, R., White, B., Lottridge, D. (2018).
[50] Forgas, J. P. (1989).

3. USERS' ACCEPTANCE OF RSs FROM SMART SPEAKERS

3.1. RESEARCH QUESTION

Many researches focus on the usage behaviours and the willingness to adopt the smart speakers. Likewise, a wide range of researches on recommender systems attempts to understand which algorithm achieve the best performances, while the best performances themselves have been identified along many dimensions (accuracy, variety, etc.). However, there are not studies concerning the user acceptance of recommender systems from smart speakers. The question is critical since many companies have adopted a successful business model based on recommender systems (Amazon, Netflix, etc.). In this stage of market expansion, a proper understanding of user acceptance could become a crucial competitive advantage for the companies which hope to extract customers' value equipping their apps for smart speaker with a voice recommender system. In order to achieve a comprehensive understanding of user acceptance of such software, many variables and their relationships must be studied.

Thus, the research questions of my study could be formulated as follows: *Which are the main variables that affect user acceptance of recommender systems of smart speakers*? *Which are their relationships*? *How can firms behave in order to receive the best response in term of acceptance of such software equipped to smart speakers*?

3.2. THEORETICAL FRAMEWORK

The acceptance of new technologies has been deeply studied for many decades and nowadays a lot of theoretical models receives high degrees of corroboration in a wide range of circumstances[51]. One of the most corroborated models is the Technology Acceptance Model (TAM), developed by Davis in 1989[52]. He identified the technology acceptance as the result of the relationship of three main variables: perceived ease of use, perceived usefulness and attitude.

The Figure 11 displays Davis' TAM[53].

[51] Sohn, K., & Kwon, O. (2020).
[52] Davis, F. D. (1989).
[53] Kim, Y. J., Chun, J. U., & Song, J. (2009).

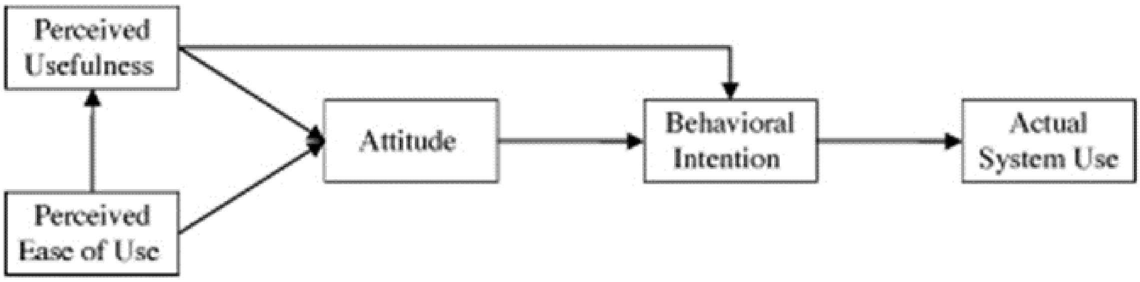

Figure 11: Original TAM (Davis 1989).

This model has been chosen because it was corroborated by many studies and it is considered a robust acceptance model. Moreover, it was adopted in studies concerning both recommender systems[54] and smart speakers[55].

To enrich the model, other variables are been added to the TAM in order to obtain a comprehensive framework. Enjoyment, Trust and Social Influence are considered important factors that affect user acceptance of recommender systems from smart speakers.

3.2.1. Perceived Ease of Use and Perceived Usefulness

Here, the Perceived Ease of Use (PEoU) is defined as "the degree to which a person believes that using a particular system is free of effort"[56]. The Perceived Usefulness (PU) is the degree to which a person believes a particular system would enhance his or her performance. People are reluctant to adopt a technology that is too difficult to use because they could perceive that the costs of use are higher than the benefits. Therefore, the PEoU is supposed to positively affect the Perceived Usefulness.

H_1: *The Perceived Ease of Use of the smart speakers positively affects the Perceived Usefulness of the recommender system of a smart speaker.*

Rogers[57] suggested, since the 60's, that the "characteristics of the innovation, as perceived by individuals, help to explain their different rate of adoption" (p. 15). One of these characteristics is the *complexity*, "the degree to which a technology is perceived as difficult to understand and use" (p. 16). Here, the difficult to use the smart speaker could become a hurdle to the adoption and acceptance of such devices. Therefore, it is important to study the PEoU as a driver of acceptance.

[54] Pu, P., Chen, L., & Hu, R. (2011).
[55] Chu, L. (2019).
[56] Davis, F. D. (1989).
[57] Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.

The PU and the PEoU could be considered as the *functional* aspects of the acceptance of new technologic systems. The main difference between the two constructs is that the PEoU mostly concerns the perceived *cost* of the use, while the PU is firstly related to the perceived *benefits*. We can expect that higher degrees of acceptance results from lower cost perception and, by opposite, higher benefit perception.

3.2.2. Attitude

The attitude is a comprehensive judgement or opinion toward an object. Here, the object of interest is the recommender system of a smart speaker. According to the TAM, Attitude (AT) is affected by both Perceived Usefulness and Perceived Ease of use.

> H_2: *Perceived Ease of Use positively affect the Attitude of a recommender system from smart speakers.*
>
> H_3: *Perceived Usefulness positively affect the Attitude of a recommender system from smart speakers.*

The main concept underlying H_2 and H_3 is that positive evaluation of benefits and costs related to the system usage results in a positive overall judgement of the recommender system itself.

3.2.3. Acceptance

The acceptance of recommender systems from smart speakers concerns the intention to use the systems. The TAM predicts a positive relationship between the PU and the Acceptance (AC), and, according to other models like the Theory of Planned Behaviour (TPB) by Ajzen (1991)[58], between Attitude and Acceptance. These relationships are supported by a plethora of research studies, even in the field of the IoT[59].

> H_4: *Perceived Usefulness positively affect the Acceptance of recommender systems from smart speakers.*
>
> H_5: *Attitude toward the recommender system of the smart speakers positively affects the Acceptance of such software.*

[58] Ajzen, I. (1991).
[59] Gao, L., & Bai, X. (2014).

Although many recent studies found a poor relationship between the attitude and the acceptance of new technologies, a study by Kim *et al*. (2009)[60] revealed that "regardless of the strength of the attitude toward using the system, attitude toward using the system is the most important determinant of behavioural intention to use the system". Similar implications were supported by other studies, such as Robinson *et al.* (2005)[61] and Chu (2019)[62].

3.2.4. Social Influence

Since the environment is risky and imperfectly predictable, people usually grasp the opinion of their social network when face the decision to buy or use a product. There is broad evidence that Social Influence (SI) affects human behaviour. Here, the Social Influence is defined as "the degree to which an individual perceives that important others believe he or she should use the new system"[63]. The "most important others" could be parents, other family components, friends, colleagues and, in a broadly perspective, everybody who is able to affect his/her behaviour, like the so called "influencers". The UTAUT model, developed by Venkatesh *et al.* (2003), suggests that the Social Influence (Venkatesh called this construct *subjective norms*) positively affects the Acceptance.

H_6: *Social Influence positively affects the Acceptance of recommender systems from smart speakers.*

As Raghunathan and Corfman (2006)[64] suggest, we can expect that "experiences in which people obtain congruent social information will be reassuring and, thus, more enjoyable, whereas those in which people obtain incongruent information will be disconcerting and, thus, less enjoyable". Similar findings result in other studies, where the SI positively affects the Enjoyment of using a system[65]. Likewise, Social Influence should have a positive relationship with Trust[66]. Indeed, according to Chaouali *et al.* (2016)[67] "an individual who believes that important others (e.g., family and friends) approve his usage of new products/services will be more inclined to trust and use these products and services". Paraphrasing Venkatesh and Davis (2000)[68], if important people believe that he or she

[60] Kim, Y. J., Chun, J. U., & Song, J. (2009).
[61] Robinson Jr, L., Marshall, G. W., & Stamps, M. B. (2005).
[62] Chu, L. (2019).
[63] Venkatesh, V., Morris, M. G., Davis, G., & Davis, F. D. (2003).
[64] Raghunathan, R., & Corfman, K. (2006).
[65] Qiu, L., & Benbasat, I. (2009).
[66] Hassanein, K. S., & Head, M. (2004, October).
[67] Chaouali, W., Yahia, I. B., & Souiden, N. (2016).
[68] Venkatesh, V., & Davis, F. D. (2000).

should perform a behaviour (e.g. using a recommender system from smart speaker), then performing it will tend to elevate his or her standing within the group. Therefore, we can assume the following hypotheses:

> H_7: *Social Influence positively affects the Enjoyment of recommender systems from smart speakers.*
> H_8: *Social Influence positively affects the Trust of recommender systems from smart speakers.*

The SI could also increase the PU of a technologic product[69] [70]. Venkatesh and Davis (2000)[71] highlight that SI "can influence intention indirectly through perceived usefulness" employing the mechanisms of compliance, internalization and identification. Under identification, others' opinions are adopted when they are associated with positively evaluated referents. Typically, "most important others" are favourable referents, so identification may reflect their opinions' influence on intent[72], affecting both the Acceptance and the PU. The compliance mechanism includes the expected "rewards" that results from the accomplishment of the social norms, imposed by the smart speaker user's reference group. With respect to internalization, it transpires when an individual consciously or unconsciously embraces others' opinions and acts in harmony with them[73]. According to the explained reasons, we should account that the SI positively affects the PU.

> H_9: *Social Influence positively affects the Perceived Usefulness of the recommender system from a smart speaker.*

3.2.5. Enjoyment

Here, the Enjoyment (E) is defined as "the extent to which the activity of using a specific system is perceived to be enjoyable in its own right, aside from any performance consequences resulting from system use"[74]. Since the *functional* performance do not account on the evaluation of E, it could be considered a *hedonic* characteristic of the system. Many studies consider the Enjoyment a strong

---

[69] Hassanein, K. S., & Head, M. (2004, October).
[70] Hassanein, K., & Head, M. (2007).
[71] Venkatesh, V., & Davis, F. D. (2000).
[72] Warshaw, P. R. (1980).
[73] Chaouali, W., Yahia, I. B., & Souiden, N. (2016).
[74] Park, Y., Son, H., Kim, C. (2012).

antecedent of technology acceptance and intention to adopt, as in Kowalczuk P. (2018)[75] concerning smart speakers acceptance. The relationship between Enjoyment and Perceived Usefulness was studied in researches concerning smart speakers and IoT devices (Sohn *et al*. 2020)[76], e-learning platforms[77] and information systems[78], where a positive effect was revealed. This could be explained by the fact that the users appreciate both the functional and the hedonic aspects of a new technology and they feel more inclined to embrace it when it is perceived as playful and funny.

> H_{10}: *Enjoyment positively affects the Perceived Usefulness of recommender systems from smart speakers.*

There are also studies concerning online shopping[79] that show a positive relationship between the Enjoyment and the Trust. Cyr *et al.* (2006)[80] argued that "enjoyment may have at least as large an impact on loyalty as perceived usefulness" and loyalty is strictly related to Trust. Meanwhile, hedonic arousal can result in a decreasing perception of the risk[81], which a strong linkage with Trust assessment.

> H_{11}: *Enjoyment positively affects the Trust toward recommender systems from smart speakers.*

The relationship between the Enjoyment and the Attitude was broadly studied in many researches. The Enjoyment is considered one of the stronger drivers of the adoption of new technologies, as in mobile services[82] and mobile recommender systems[83]. Utilitarian and hedonic aspects both shape Attitude and were considered in several model of behaviour prediction, as Ahtola (1985)[84] noticed: "the utilitarian aspect is considered to be separate from the hedonic aspect, not a subconstruct of it" and "the utilitarian and hedonic aspects of attitudes together determine the third attitudinal construct that I call the general aspect of an attitude", which is the one generally accepted. According to

[75] Kowalczuk, P. (2018).
[76] Sohn, K., & Kwon, O. (2020).
[77] Abdullah, F., Ward, R., & Ahmed, E. (2016).
[78] Qiu, L., & Benbasat, I. (2009).
[79] Hwang, Y., & Kim, D. J. (2007).
[80] Cyr, D., Head, M., & Ivanov, A. (2006).
[81] Sarkar, A. (2011).
[82] Nysveen, H., Pedersen, P. E., & Thorbjørnsen, H. (2005).
[83] Choi, J., Lee, H. J., Sajjad, F., & Lee, H. (2014).
[84] Ahtola, O. T. (1985).

Childers *et al.* (2001)[85] Enjoyment is a strong predictor of Attitude, together to the PEoU and the PU in online web context.

> H_{12}: *Enjoyment positively affects the Attitude toward recommender systems from smart speakers.*

Citing the researches of Moon and Kim (2001)[86] and Swanson (1987)[87], Hassanein and Head (2004) said that "systems that are easier to use will be less threatening and will encourage feelings of control. This, in turn, can result in a more enjoyable experience with the technology."[88] Then, we can account that Perceived Ease of Use would positively affects Enjoyment.

> H_{13}: *Perceived Ease of use of smart speakers positively affects Enjoyment.*

3.2.6. Trust

Here, the Trust indicates the degree of which the recommender systems from smart speakers are found trustworthy and they are in users' best interests. Concerning recommender systems evaluation, Pu *et al.* (2011) found a positive effect of Perceived Usefulness on Trust[89]. The same effect was found by Suh and Han (2002) on their research about consumers' acceptance of internet banking[90] and by Ingham and Cadieux (2016) about e-shopping[91] [92]. This last study identifies also a positive effect of Perceived Ease of Use on Trust, which was shown also in the adoption of the electronic logistics information systems[93].

> H_{14}: *Perceived Usefulness positively affects Trust toward recommender systems from smart speakers.*
> H_{15}: *Perceived Ease of Use positively affects Trust toward recommender systems from smart speakers.*

---

85 Childers, T. L., Carr, C. L., Peck, J., & Carson, S. (2001).
86 Moon, J. W. & Kim, Y. G. (2001).
87 Swanson, E. B. (1987).
88 Hassanein, K. S., & Head, M. (2004, October).
89 Pu, P., Chen, L., & Hu, R. (2011).
90 Suh, B., & Han, I. (2002).
91 Ingham, J., & Cadieux, J. (2016, January).
92 Gefen, D., Karahanna, E., & Straub, D. W. (2003).
93 Tung, F., Chang, S., & Chou, C. (2008).

The relationship between Trust and Attitude was received a broad degree of corroboration in many studies, about acceptance of personalized business models[94], social presence through web interfaces[95], online shopping[96] and trading[97].

H_{16}: *Trust positively affects the Attitude toward recommender systems from smart speakers.*

The resulting conceptual model is exhibited in the Figure 12.

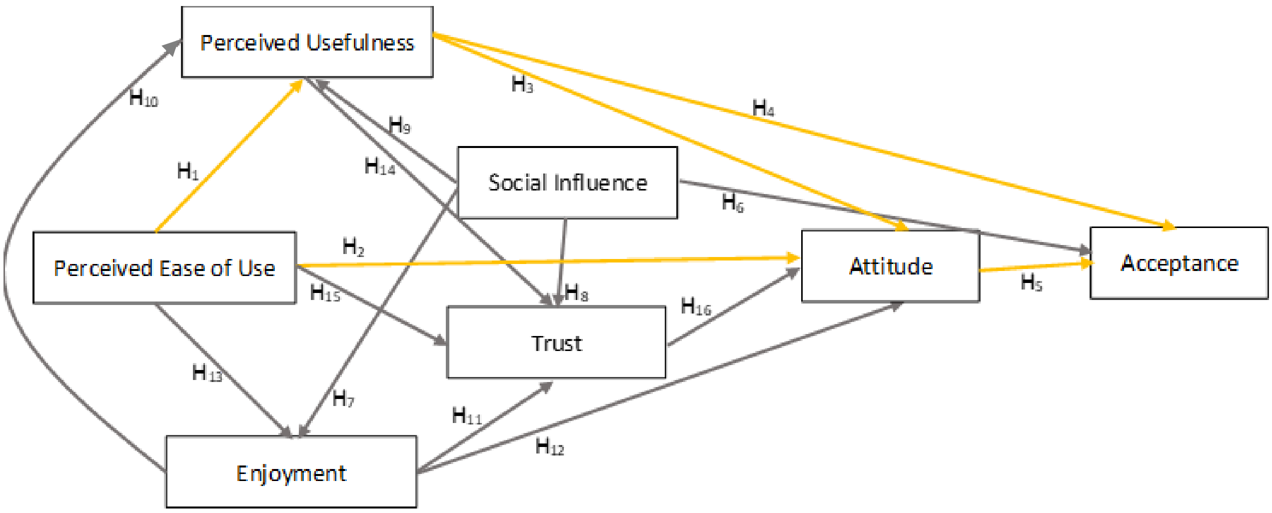

Figure 12: Conceptual model. In yellow, the original TAM hypotheses.

3.3. RESEARCH DESIGN

In order to answer the research questions, a survey questionnaire was developed (the questions are available in the Appendix to this work). The scale variables were gathered in 5-point likert scales (1 – completely disagree; 5 – completely agree), then some demographics were asked (gender, age and employment). First at all, it was asked whether the respondent knows what a smart speaker is. If he/she responds "No", a brief description of the devices was provided, together to the image of both Google Home and Amazon Echo. Then, the questionnaire goes on the same for both groups of respondents. The questionnaire was undertaken to my contacts and posted in university social networks to reach as many respondents as possible, during the period between 4th March and 13th

[94] Zhao, J., Fang, S., & Jin, P. (2018).
[95] Hassanein, K., & Head, M. (2007).
[96] Gefen, D., Karahanna, E., & Straub, D. W. (2003).
[97] Lee, M. (2009).

March 2020. In total, 286 responses were collected. The average age of the respondents is 23.96 years (standard deviation 5.69, 76 missing values), while the sample is made up 141 females (67%, 75 missing values). The 74% (212) of the respondents completed the entire survey, one of those missed only the demographic box. The respondents are mainly students (84%, 76 missing values).

3.4. DATA ANALYSIS

Only the 212 respondents who completed the scales questions were included in the data analysis. In order to clean the data from not reliable responses, it was checked if there were records with same responses in all the fields. No respondents were delated in this way for further analyses.

96% of the respondents says they know what a smart speaker is, while the remaining 9 respondents says they are not aware of what a smart speaker is. 111 respondents (52%) says they have used a smart speaker at least once, while only 56 own a smart speaker (26%). 8 (14%) of them own such device for less than one month; 18 (32%) from one to six months; 14 (25%) from six months to one year; other 14 (25%) from one to two years; the remaining 2 (4%) for more than two years. The charts below exhibit the usage and ownership data collected by the survey.

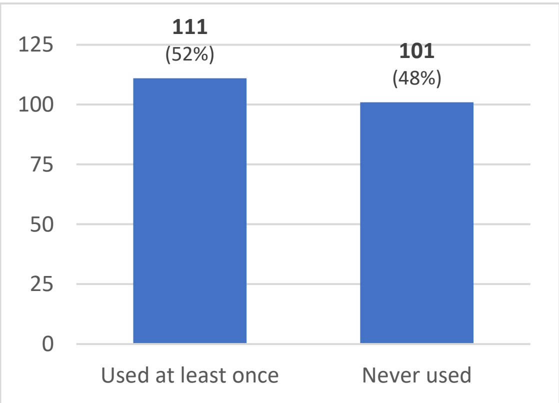

Figure 13: Have you ever used a smart speaker? (sample size: 212)

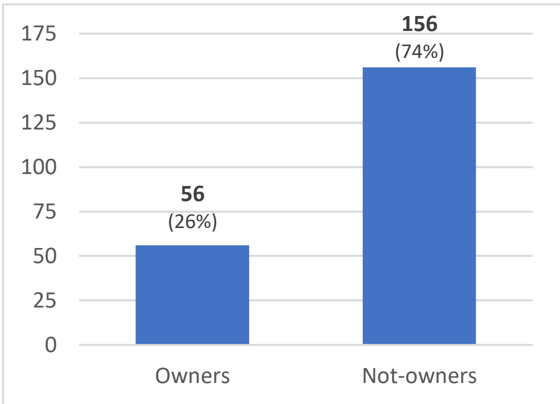

Figure 14: Do you own a smart speaker? (sample size: 212)

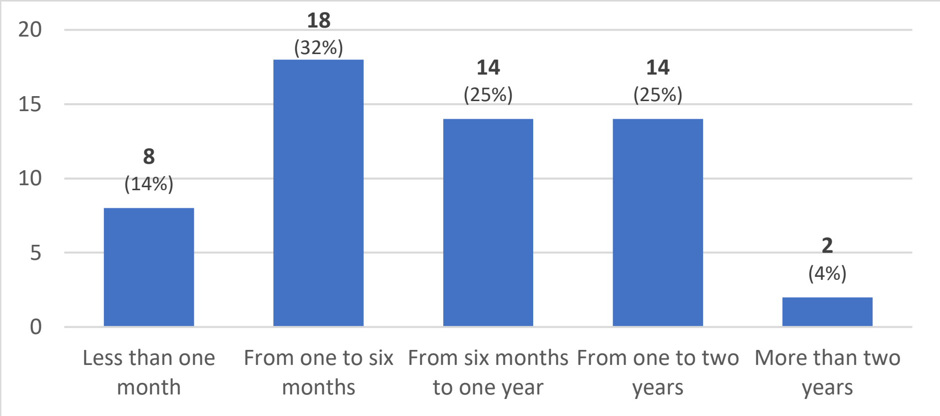

Figure 15: How long have you owned a smart speaker?
(on 56 respondents who own a smart speaker)

The questions refer to the offering to listen to a song made by the recommender system from a smart speaker. The music was chosen because both males and females feel comfortable with it and there is corroborated evidence that the smart speakers are mostly used to listen to music[98]. Newman (2018) finds that music is both the most valued and regularly used features in UK, quoting that “ the ability to play music comes out on top, with four in five (84%) saying they use this feature and almost two-thirds (61%) saying it is their most valued feature”.

The scale items were tested for both internal reliability and external validity, performing Cronbach alpha and principal component analysis. The scale items were changed to obtain acceptable values of validity and reliability. The reverse questions were removed, while the item AT_1 (from Attitude scale) was attributed to Enjoyment scale. Even the T_3 item was removed from Trust scale. In the table below, the final Cronbach alpha and factor loadings of each scale item are displayed.

	ALPHA	LOADING	SCALE
PEOU_1	0.68	0.83	Perceived Ease of Use
PEOU_2	0.68	0.90	Perceived Ease of Use
PU_1	0.89	0.82	Perceived Usefulness
PU_2	0.89	0.81	Perceived Usefulness
PU_3	0.89	0.78	Perceived Usefulness
E_1	0.81	0.79	Enjoyment
E_3	0.81	0.69	Enjoyment
SI_1	0.84	0.87	Social Influence
SI_2	0.84	0.92	Social Influence
T_1	0.85	0.85	Trust
T_2	0.85	0.87	Trust
AT_1	0.81	0.52	Enjoyment
AT_3		0.56	Attitude
AC_1	0.86	0.82	Acceptance
AC_2	0.86	0.81	Acceptance

Table 3: For each scale item, it is exhibited the Cronbach alpha, the factor loading and the scale which belongs to.

98 Bentley, F., LuVogt, C., Silverman, M. Wirasinghe, R., White, B., Lottridge, D. (2018).

As Gliem & Gliem quoted (2003)[99], George & Mallery (2003)[100] provide the following rule of thumb to evaluate the Cronbach's alpha: "_ > .9 – Excellent, _ > .8 – Good, _ > .7 – Acceptable, _ > .6 – Questionable, _ > .5 – Poor, and _ < .5 – Unacceptable" (p. 231). All the scales exhibit a good internal consistency reliability, while Perceived Ease of Use is questionable. The external validity was tested by factor loadings. We use the Hair *et al.*'s (1998) criterion of assessing variables to constructs: "variables with loading greater than 0.3 were considered significant; loading greater than 0.4, more important; and loadings of 0.5 or greater were quite significant."[101]

Descriptive statistics for each scale are provided in the following table.

SCALE	N. OF ITEMS	MEAN	STD DEVIATION
Perceived Ease of Use	2	3.85	0.87
Perceived Usefulness	3	3.33	1.08
Enjoyment	3	3.18	0.98
Social Influence	2	1.93	0.91
Trust	2	2.89	0.98
Attitude	1	2.84	1.13
Acceptance	2	3.48	1.03

Table 4: Scale descriptive statistics (5-point likert scales).

The highest values are exhibited by PEoU scale, meaning that the use of smart speakers is perceived effortless. The SI scale is in average very low, meaning that the people who influence the respondents' behaviours do not oppose the use of smart speakers. Therefore, the respondents do not experience a feeling of social pressure on the intention to use recommender systems from smart speakers.

Here, independent t-tests were performed among males and females in order to discover whether the two groups differ along the scales. It results that there are not significant differences between male and female in any scale construct. Higher Perceived Usefulness is exhibited by males (mean = 3.48) with respect to the females (mean = 3.20), however it is significant at the 90% confidence interval ($p < 0.10$), therefore it could be considered poorly significant.

The acceptance was even tested on the two groups of the people who have never used a smart speaker and the ones who have used the smart speakers at least once. The independent t-test reveals that who have used a smart speaker at least once exhibit a higher acceptance of recommender systems from smart speakers (mean = 3.61) than who have never interacted with such devices (mean = 3.34),

[99] Gliem, J. A., & Gliem, R. R. (2003).
[100] George, D., & Mallery, P. (2003).
[101] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998).

although the difference is poorly significant ($p < 0.10$). The Table 5 shows the mean of the responses, split by gender and "used" (respondents who have used a smart speaker at least once) against "never-used" (respondents who have never used a smart speaker), with their relative p-values.

SCALE	GROUPS	MEAN	T-STATISTIC	DF	P-VALUE
Perceived Ease of Use	Males	3.91	0.66	209	0.51
	Females	3.82			
	Used	3.92	1.18	210	0.24
	Never used	3.78			
Perceived Usefulness	Males	3.48	1.81	209	0.07
	Females	3.20			
	Used	3.38	0.88	210	0.22
	Never used	3.19			
Enjoyment	Males	3.20	0.29	209	0.77
	Females	3.16			
	Used	3.29	1.73	210	0.09
	Never used	3.05			
Social Influence	Males	2.02	1.24	209	0.21
	Females	1.86			
	Used	2.09	2.73	210	0.01
	Never used	1.75			
Trust	Males	2.97	0.90	209	0.37
	Females	2.84			
	Used	3.01	1.81	210	0.07
	Never used	2.77			
Attitude	Males	2.87	0.25	209	0.80
	Females	2.83			
	Used	2.97	1.55	209	0.12
	Never used	2.72			
Acceptance	Males	3.44	0.38	209	0.70
	Females	3.49			
	Used	3.61	1.89	209	0.06
	Never used	3.34			

Table 5: t-test on males/females and used/never-used. Means of each group are provided, together to the t-statistic, the degrees of freedom (df) and the p-value. Only the Social Influence between the used and never-used groups is statistically significant at 95% ($\alpha = 0.05$), where the p-value = 0.01.

The correlations between the scales are computed and the p-values are tested before further analyses. The Table 6 contains the Pearson's correlations between the scales.

	PEoU	PU	E	SI	T	AT	AC
PEoU	1.00***						
PU	0.12*	1.00***					
E	0.08	0.77***	1.00***				
SI	0.09	0.38***	0.40***	1.00***			
T	0.17**	0.49***	0.53***	0.31***	1.00***		
AT	0.05	0.65***	0.67***	0.35***	0.54***	1.00***	
AC	0.19***	0.62***	0.69***	0.26***	0.54***	0.56***	1.00***

*Table 6: The Pearson's correlations between the scales are exhibited in the table. * p-value<0.1; ** p-value<0.05; *** p-value<0.01.*

The PEoU exhibits the poorest correlations, even not significant ones; while the correlations between the other variables are all significant with a p-value lower the 0.01. The stronger correlation is between Perceived Usefulness and Enjoyment (0.77), followed by the correlation between E and Acceptance (0.69), E and Attitude (0.67), PU and AT (0.65) and between PU and AC (0.62). The Perceived Usefulness and the Enjoyment would play a crucial role on Acceptance of recommender systems from smart speakers because of their high correlation with AC. However, the structural equation modelling will help us to reach a deeper understanding of these results.

Here, Structural Equation Modeling (SEM) was performed to test the hypotheses. The open source R environment was employed, using the “lavaan” package from Rosseel (2012)[102]. It performs an iterative method, involving the definition of a set of *start values* of the free parameters. These default values are corrected at each iteration comparing the resulting covariance matrix with the observed (actual) covariance matrix. The iterations end when no change occurs (or a tolerance difference is reached).

The so called *measurement model* is made of observed variables (the scales items), while the *structural model* is the one showed in Figure 12 (the latent variables).

The measurement model employs the following equations:

$$PEoU = \sim PEoU_1 + PEoU_2$$
$$PU = \sim PU_1 + PU_2 + PU_3$$
$$E = \sim E_1 + E_3 + AT_1$$
$$SI = \sim SI_1 + SI_2$$
$$T = \sim T_1 + T_2$$
$$AC = \sim AC_1 + AC_2$$

The regressions in the structural model are formulated as follows:

[102] Rosseel, Y. (2012).

$$E \sim PEoU + SI$$

$$PU \sim PEoU + SI + E$$

$$T \sim PEoU + SI + E + PU$$

$$AT_3 \sim PEoU + E + PU + T$$

$$AC \sim PU + SI + AT_3$$

We employed the same syntax of Rosseel, where the "=~" operator means *"is manifested by"* and indicates the manifest measures which build the latent construct. The tilde sign "~" is the regression operator. In the left-side of the formula there is the independent variable (y), while in the right-side there are the dependent variables (x_i), which explain the independent variable. The result of the SEM analysis is a kind of network of regression analyses, where the variables are built as factors of the manifest variables. However, it is not the same to perform in a two-stage analysing involving a factor analysis followed by a regression analysis.

The result of the SEM is displayed in Figure 16.

Regressions:	Estimate	Std.Err	z-value	P(>|z|)
E ~				
PEoU	0.026	0.044	0.595	0.552
SP	0.370	0.074	5.028	0.000
PU ~				
PEoU	0.016	0.028	0.585	0.559
SP	-0.005	0.047	-0.097	0.923
E	0.989	0.087	11.378	0.000
TT ~				
PEoU	0.037	0.062	0.598	0.550
SP	0.027	0.066	0.403	0.687
E	0.672	0.305	2.202	0.028
PU	0.028	0.269	0.104	0.917
AT_3 ~				
PEoU	-0.000	0.011	-0.022	0.982
E	0.702	0.273	2.573	0.010
PU	0.067	0.231	0.289	0.773
TT	0.222	0.082	2.719	0.007
AC ~				
PU	0.627	0.092	6.807	0.000
SP	0.018	0.052	0.347	0.729
AT_3	0.132	0.067	1.956	0.051

Figure 16: The result of the SEM. Standardized β coefficients are in the first column, while in the last one there are the p-values related to each tested relationship.

The Perceived Ease of Use has not any significant effect on the other variables (H_1, H_2, H_{13} and H_{15} are rejected). Even the Social Influence has not significant effect on Perceived Usefulness, Trust and Acceptance (H_6, H_8 and H_9 are rejected). Perceived Usefulness has not significant effect on Trust and Attitude (H_3 and H_{14} are rejected). Table 7 summarizes the supported and unsupported hypotheses, with their respective p-values and β coefficients.

N.	HYPOTHESIS	β-COEFFICIENT	P-VALUE	EVALUATION
H_1	PEoU $\rightarrow$ PU	0.02	0.56	rejected
H_2	PEoU $\rightarrow$ AT	-0.00	0.98	rejected
H_3	PU $\rightarrow$ AT	0.07	0.77	rejected
H_4	PU $\rightarrow$ AC	0.63	< 0.01	supported
H_5	AT $\rightarrow$ AC	0.13	< 0.05	poorly supported
H_6	SI $\rightarrow$ AC	0.02	0.73	rejected
H_7	SI $\rightarrow$ E	0.37	< 0.01	supported
H_8	SI $\rightarrow$ T	0.03	0.69	rejected
H_9	SI $\rightarrow$ PU	-0.01	0.92	rejected
H_{10}	E $\rightarrow$ PU	0.99	< 0.01	supported
H_{11}	E $\rightarrow$ T	0.67	< 0.05	poorly supported
H_{12}	E $\rightarrow$ AT	0.70	< 0.01	supported
H_{13}	PEoU $\rightarrow$ E	0.03	0.55	rejected
H_{14}	PU $\rightarrow$ T	0.03	0.92	rejected
H_{15}	PEoU $\rightarrow$ T	0.04	0.55	rejected
H_{16}	T $\rightarrow$ AT	0.22	< 0.01	supported

Table 7: original model hypotheses evaluation, β coefficients and p-values.

Hence, another SEM analysis was performed only on significant relationships in order to obtain a clearer representation of the network. The below Figure 17 displays the results.

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  E ~
    SP                0.375    0.073    5.136    0.000
  PU ~
    E                 0.998    0.079   12.663    0.000
  TT ~
    E                 0.737    0.085    8.717    0.000
  AT_3 ~
    E                 0.784    0.097    8.110    0.000
    TT                0.208    0.078    2.686    0.007
  AC ~
    PU                0.625    0.088    7.132    0.000
    AT_3              0.141    0.066    2.124    0.034
```

Figure 17: Results of SEM on final model.

All the relationships are significant ($p < 0.01$), except the relationship between Attitude and Acceptance that exhibits a poorer significance than the others ($p < 0.05$). Acceptance of recommender systems from smart speakers is mainly explained by the Perceived Usefulness and (indirectly) by Enjoyment, while the Attitude seems to have a poorer effect. Many studies discover the same result, where AT poorly or not at all affects the AC. Consistent with many other research studies, even the

SI appears to have not direct effect on AC, such as in Chaouali *et al.* (2016)[103]. At the opposite, the Enjoyment appears to have a central role on the model of acceptance of recommender systems from smart speakers, since it replaces the role of PEoU of original TAM. It has strong direct effects on Trust ($\beta = 0.74$, $p < 0.01$), Attitude ($\beta = 0.78$, $p < 0.01$) and Perceived Usefulness ($\beta = 1.00$, $p < 0.01$). It exhibits also significant indirect effect on AT, when mediated by T ($\beta = 0.15$, $p < 0.01$), and on AC, when mediated by PU ($\beta = 0.63$, $p < 0.01$). The Table 8 summarizes the beta coefficients and the significance level of the respective hypotheses of the adjusted model.

N.	HYPOTHESIS	β-COEFFICIENT	P-VALUE	EVALUATION
H_4	PU → AC	0.63	< 0.01	supported
H_5	AT → AC	0.14	< 0.05	poorly supported
H_7	SI → E	0.38	< 0.01	supported
H_{10}	E → PU	1.00	< 0.01	supported
H_{11}	E → T	0.74	< 0.01	supported
H_{12}	E → AT	0.78	< 0.01	supported
H_{16}	T → AT	0.21	< 0.01	supported

Table 8: final model hypotheses evaluation, β coefficients and p-values.

Then, the R^2 for each scale are provided in Table 9.

Scale	E	PU	T	AT	AC
R^2	0.22	0.84	0.40	0.58	0.58

Table 9: R^2 for each scale are provided.

R^2 represents the proportion of the variance of each scale explained by the model. The Enjoyment scale is the one which exhibited the poorest R^2, where the Social Influence explains only the 22% of Enjoyment total variance. The PU is the scale that has the highest R^2, where the E is able to explain the 84% of PU total variance.

Here, the overall model fit was evaluated using some statistics which are commonly employed in SEM: the Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Goodness-of-Fit Index (GFI), Normed Fit Index (NFI), Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean square Residual (SRMR). The $\chi2$ (chi-square) statistic tests the overall model fit. $\chi2$ is 170 with 59 degrees of freedom (df) and $p < 0.001$. $\chi2/df$ is 2.88 (a value lower than 5 or around 2 are considered good)[104]. Hooper *et al.* collected the suggestions from previous research studies for the cut off criteria of the fit indexes. In general, RMSEA between 0.10 and 0.08 provides mediocre

[103] Chaouali, W., Yahia, I. B., & Souiden, N. (2016).
[104] Hooper, D., Coughlan, J., & Mullen, M. R. (2008).

fit, while values under 0.08 provide good fit. SRMR less the 0.08 are considered fair fit, while values less than 0.05 are considered good. TLI, GFI, CFI and NFI higher than 0.90 indicates well-fitting models, while in some circumstances it was indicated a higher threshold of TLI, CFI and NFI on 0.95. Therefore, we can consider the overall fit of the model acceptable, considering also that with a cut off value of 0.95 there was a slight tendency for TLI and CFI to over-reject true-population models at small sample size (N < 250)[105].

χ2/df	TLI	CFI	GFI	NFI	RMSEA	SRMR
2.88	0.92	0.94	0.89	0.91	0.09	0.05

Table 10: The goodness of fit indexes of the final model

The final model is exhibited in Figure 18.

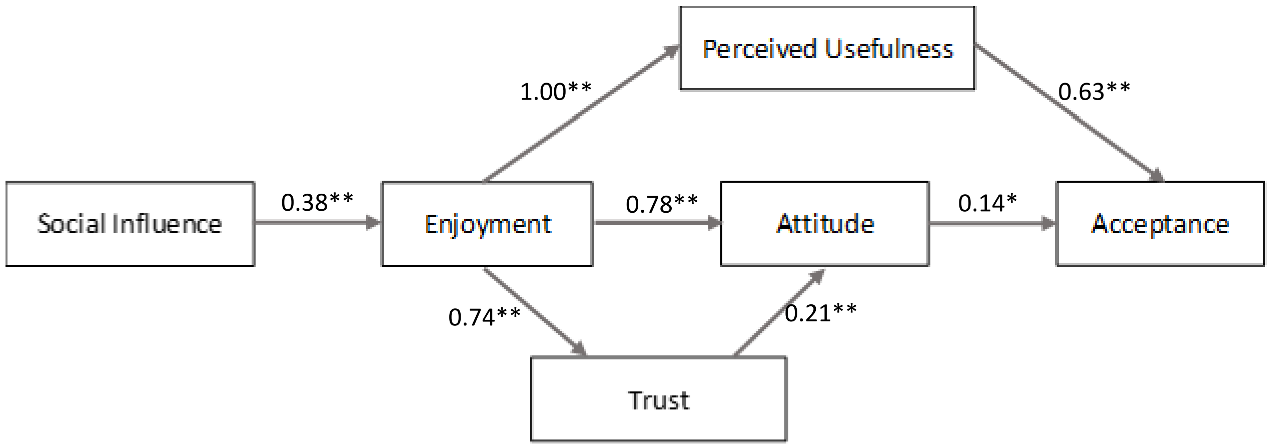

*Figure 18: Final model of user acceptance of recommender systems from smart speakers. β coefficients are represented on the arrows. * p-value<0.05; ** p-value<0.01.*

3.5. DISCUSSION

On average, people how used the smart speakers at least once show greatest perceived social approval from the use (highest SI scores) and tend to have higher acceptance of recommender systems from smart speakers then the people who have never used such devices. This behaviour should be interpreted on the light that even the adoption of the smart speakers and, in general, the IoT and artificial intelligence-based products is affected by this driver.

No significant differences are found in the males' responses with respect to the females' ones. Only the Perceived Usefulness of the recommender systems from smart speakers seems to exhibit a poorly significant higher score in males than in females' group. Nevertheless, the gender does not significantly affect the Acceptance of such systems.

[105] Hu, L. T., & Bentler, P. M. (1999).

The SEM analysis enables us to have a deep understanding of the relationships between the variables which affects the acceptance of the recommender systems from smart speakers. We built a model based on Davis' TAM and enriched it introducing the Trust, the Social Influence and the Enjoyment, resulting in a new acceptance model. Most of the hypotheses were rejected due to the results of the SEM, however crucial acknowledges were reached. The Perceived Ease of Use, which is a salient variable in TAM, does not exhibit any significant effect on other variables of the model. The place of PEoU is taken by the Enjoyment, which is also the most effective driver of acceptance of recommender systems from smart speakers. It is meaningful that E strongly affects both the PU and the Attitude, as PEoU makes in the Davis' TAM. This means that the Perceived Ease of Use does not have a prominent role on explaining the acceptance in this peculiar case. The relatively poor effort which is required to the user of smart speakers in the usage phase (it is not the same in the setup phase of the device) could explain the marginal role of PEoU. Voice commands are easy to deliver and both the people who experimented the smart speakers and the ones who have never used them perceive this ease of use. It is not by chance that PEoU scale scored the highest values (mean = 3.85 out of 5.00) among the constructs.

The prominent role of the Enjoyment on the adoption of IoT devices was recently studied by Sohn & Kwon (2020)[106] who confirm even the importance (in descending order) of Perceived Usefulness, Subjective Norms (*alias* Social Influence) and Attitude, which do not exhibit a strong relationship with acceptance, while the PEoU explains only the 2.56% of the behavioural intention to use such devices[107]. These results suggest that these systems, and the smart speakers in general, could be interpreted as hedonic services (or products), since their acceptance and perceived usefulness strictly depend on affective factors.

From a managerial perspective, it seems wise to leveraging on the perception of Enjoyment and PU in order to reach the target levels of acceptance of recommender systems from smart speakers. This means that the company communication and the product advertising should focus on the playfulness of the systems and the ways the recommender systems could help the users to enhance their efficacy.

By contrast, the Social Influence poorly affects the E, while neither no significant effect on Trust and PU nor a direct effect on acceptance were discovered. The SI exhibits also the lowest average scores, meaning that perceived social approval is not a salient issue on the acceptance of recommender systems from smart speakers.

[106] Sohn, K., & Kwon, O. (2020).

[107] The comparison between Shon and Kwon's results and our research must be intended on the voice assistant and not on smart speaker adoption, since the recommender systems from smart speakers are more related to the software side than the adoption of the hardware.

A marginal role was relegated also to Trust, which mediates the E and the AT, but its effect on Acceptance is very poor. Although security and privacy concerns affect the smart speakers, they do not result in a negative perception of the trustworthiness of the recommender systems from such devices. Many studies encourage this interpretation, since people claim to be sensitive to privacy and security issues, but in their actual behaviours these concerns have poor, or even not at all, relevance. This inconsistency of privacy attitudes and privacy behaviour is often referred to as the "privacy paradox"[108].

[108] Kokolakis, S. (2017).

4. CONSLCUSIONS

4.1. RESEARCH IMPLICATIONS

Starting from original TAM model by Davis, we developed a user acceptance model of recommender systems from smart speakers. The TAM yields to less reliable results than the model developed in this research, since the user's Perceived Ease of Use seems to have no significant effect on any of the other variables, even on Acceptance.

The crucial role of Enjoyment emerges from a plethora of previous researches on usage and acceptance of smart speakers, and it was confirmed even on acceptance of recommendation systems from such devices. Kowalczuc (2018)[109] reveals that Enjoyment is the strongest driver of Acceptance of smart speakers, followed by Perceived Usefulness and Perceived Risk (this construct considers both the security and privacy concerns about these devices), while the Perceived Ease of Use has no significant effect on behavioural intention. Likewise, Sohn & Know (2020)[110] demonstrated that Enjoyments explains around 33% of behavioural intention to use the voice assistants (such as Siri, Cortana, or Alexa), while it is followed by PU that accounts for only the 18%.

Here, we can confront the results of our study to Sohn & Know's ones, revealing strong similarities (Figure 19).

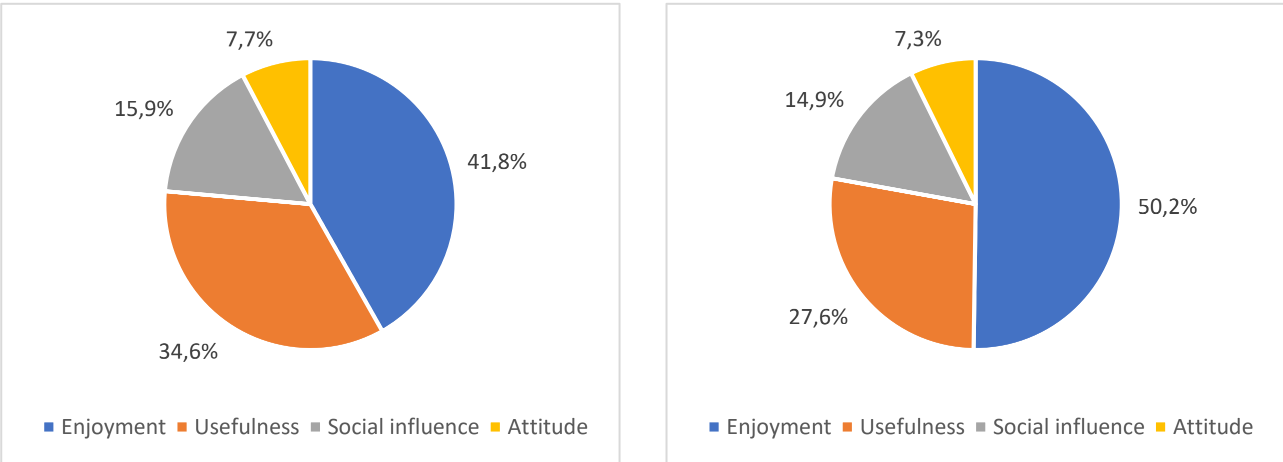

Figure 19: the proportion of the influence of E, PU, SI and AT on Acceptance. Only the comparable variables between this study (on the left) and Sohn & Know's one (on the right) are considered. On the right, our reworking of Sohn & Know's findings.

[109] Kowalczuk, P. (2018).
[110] Sohn, K., & Kwon, O. (2020).

The Enjoyment role as a strong driver of acceptance of smart speakers is demonstrated not only on users' intention, but even on actual usage behaviours. The voice command "tell me a joke" is emblematic to explain the pervasive role of E on smart speakers. Bentley *et al.* (2018)[111] analysed voice history logs of 65,499 interactions with Google Home and found that the command "tell me a joke" was uttered in 0.8% of cases. It could seem a poor result, but we have to considered that only eight other commands overcome it: "stop" (7.3%); "what time is it" (2.7%); "pause" (1.1%); "how much time is left" (1.1%); "pause TV" (0.8%); "play" (0.8%); "skip this song" (0.8%); "volume up" (0.8%). Some of these commands are routinely tasks, like "stop", "play" and "pause"; two concerns the time and the other two the volume and the songs.

Sciuto *et al.* (2018)[112] studied 278,654 commands, finding that the "tell me a joke" accounts for around 1.0% of the total, a percentage comparable with the "tell me the news" command (1.1%).

Then, we can be pretty confident that the Enjoyment is a crucial variable that the marketers could exploited when advertise from smart speakers or when they built recommender systems for such devices. Funny and playful ways to recommend items seem to strongly affect the acceptance, and then, the adoption and the persuasiveness of RS in this field.

Another well-corroborated driver of acceptance is the Perceived Usefulness (PU) of smart speakers, in general, and of RS for such devices, in particular. Recommendation systems are known to be an effective way to cope with the information overload issue. The marketers should leverage on the possibility to have more personalized results when looking for offerings to purchase and on the effectiveness of the recommendations which helps the user to save time and mental effort on their researches. Actually, the researches could be time-consuming and the unsatisfied consumers are annoyed and frustrated when they do not achieve their purposes, such as finding a place to go on holiday, or choosing a movie to see during the night. Even the user's attention decreases when the time spent on the research rises. Therefore, the marketers must teach to the users that the functionalities of the RS are on users' best interests and they could be useful to enhance user's efficacy and effectiveness of the research.

Moreover, the performance of the RSs must be evaluated both by accuracy and serendipity measures (paragraphs 2.3). The accuracy reflects the functional aspect of recommender systems and it concerns the degree by which the recommended items reflects the actual users' preferences, then it measures the usefulness of the suggestions. Whereas, the serendipity measures the users' emotional responses, then it is more suitable to evaluate the playfulness and fun of the RS and the respective recommended items.

[111] Bentley, F., LuVogt, C., Silverman, M. Wirasinghe, R., White, B., Lottridge, D. (2018).
[112] Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018, June).

Hence little or no differences are been found between males and females, it is not suggested to marketers to leverage on this group distinction. Contrarily, the respondents who have used the smart speakers at least once show a significant higher acceptance than the respondents who have never used such devices. This means that the acceptance of recommender systems could be improved just encouraging the trial of such systems. Tylor & Todd (1995)[113] said that "prior experience has been found to be an important determinant of behaviour" and that "it has been suggested that knowledge gained from past behaviour will help to shape intention". Thus, the marketers must encourage trial, perhaps by decreasing the costs associated with the systems (ease of installation of the "skills", decreasing the price of smart speakers) or by offering promotions to the customers who use the RS from smart speakers.

4.2. THE ROLE OF THE TRUST

This study highlights the marginal role played by Trust (T) on the Acceptance of recommender systems from smart speakers (it explains around 1.6% of AC). The result could partially be explained by the fact that the object of the study was song recommendation, which do probably not arise risk concerns. We can except that whether the risk perception of the usage of the RS rises, even the trust would become a strong driver of acceptance.

The risk increases when the expected losses rises too, e.g. when the recommendation refers to a purchase. Since the purchase involves a fee to be payed, the trustworthiness of the system is expected to affect the adoption and acceptance. Indeed, Lai *et al.* (2012)[114] found that "trust, privacy and security and switching cost, have the strongest effect on switching intentions" toward mobile shopping.

Moreover, it was studied that the human-likeness of some characteristics of smart speakers could enhance the trust toward the device. Hu *et al.* (2019)[115] state that the people are more likely to perceive social presence and trust when they experienced matched human-likeness of speaking and listening. Therefore, it is possible to enhance Trust making the conversation between the user and the smart speaker more natural. Qui & Benhbast (2012)[116] found that humanoid embodiment and natural speech of product recommendation agents results on higher perception of social presence and trust.

However, Trust can be negatively affected by privacy and safety concerns. From the privacy point of view, the smart speakers must be listening at all time the users, even when they are not aware of

[113] Taylor, S., & Todd, P. (1995).
[114] Lai, J. Y., Debbarma, S., & Ulhas, K. R. (2012).
[115] Hu, P., Lu, Y., & Gong, Y. (2019).
[116] Qiu, L., & Benbasat, I. (2009).

it. The companies assure users that no personal audio leaves the devices until proper initialization[117], however, the smart speakers could be accidentally activated by false positive wake words, allowing the audio data collection when the users are not conscious of it. False positive wake words even affect the safety of smart speakers because it was noted that voice assistants are vulnerable to inaudible ultrasonic frequencies. An attacker could interact with a smart speaker with ultrasonic frequencies and the device responds to his/her commands. This type of attack can be also embedded in malicious advertisements, as Hoy (2018) suggests. Likewise, Holt (2018) quotes that even false positive wake words can be embedded in malicious advertisements to activate the smart speakers.

However, this study finds that trust on recommendations from smart speakers seems to be not an issue, while, perhaps, the main anxieties are on the privacy and security problems of the devices themselves, as many researches demonstrated. Namely, Wharton (2018) reports that McAfee survey of US internet users in November 2017 found "security was one of the most important features respondents look for when deciding to buy a smart-home device – more so than price or ease of use". 36% of the respondents said security was a feature they looked for before buying a smart-home device, such as a smart speaker, while 32% of them said price was the most important feature to consider before purchasing a smart-home device (just 13% cited ease of use). This is a crucial issue for the marketers who want to develop their own third-party skill as an online shop platform equipped with RSs. Marriot & Williams (2018) found that overall trust is "the most significant predictor of intention" in mobile shopping activities, and it is logical to expect that similar results would be find on online shopping via smart speakers.

4.3. LIMITATIONS AND FUTURE RESEARCHES

The research study we have conducted has some limitations that we need to underline. The first one is the poor variability of the sample of respondents who participated the experiment. Almost all the respondents are students with an age that ranges from 19 to 27. Further researches could be performed in order to discover whether the age or the social status of the users significantly affects the acceptance model we have proposed. Even the nationality could have a role on shaping the weights of the variables, while the proposed study was conducted only on Italian respondents.

The recommendation item we considered was the songs, since the music applications are the most used from smart speakers. Nevertheless, marketers are also interested on selling products, then further researches could focus on users' purchase intentions elicited by RS from smart speakers. However,

[117] Holt (2018).

shopping via smart speakers is still very unpopular. Nonetheless, eMarketer[118] reported that 15% of US smart speaker owners asked their devices for a product recommendation (18% of them are males, 15% are females) and the market is growing.

As we have already seen (paragraph 1.1), marketers could build their own third-party skills to promote their offerings. The marketers could equip these applications with recommender systems in order to take advantage from this technology. Further researches would involve the role played by the trustworthiness of the third-party skill provider on shaping the users' willingness to adopt and accept the RSs (as the previous paragraph suggested).

Another limitation of this work is represented by the Attitude (AT) scale, which is made of only one question item. Attitude is a complex construct which needs a multi-item scale to yield more reliable results. Even though, in order to maintain internal reliability and external validity of the constructs, we decided to reduce this construct to a single-item scale. Further researches could leverage on this disadvantage and produce more sophisticated measures of AT. Moreover, it would be considered to enrich the Attitude measure distinguishing the cognitive from the affective attitude. The cognitive attitude deals with the believes about the technology, while the affective one is about favourability and likeness. Yang & Yoo (2004) tested the TAM along the cognitive and affective dimensions of attitude. They administered a survey questionnaire to U.S. students and found that cognitive dimension has a significant stronger mediator effect on Acceptance than the affective one. The quoted study would suggest to looking for more complex constructs of attitude in order to reach more reliable results.

[118] Koch, L. (2019).

BIBLIOGRAPHY

1. Abdi, N., Ramokapane, K. M., Such, J. M. (2019). More than smart speakers: security and privacy perceptions of smart home personal assistants. *In Fifteenth Symposium on Usable Privacy and Security*.

2. Abdullah, F., Ward, R., Ahmed, E. (2016). Investigating the influence of the most commonly used external variables of TAM on students' Perceived Ease of Use (PEOU) and Perceived Usefulness (PU) of e-portfolios. *Computers in Human Behavior*, 63, 75-90.

3. Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS quarterly*, 665-694.

4. Ahtola, O. T. (1985). Hedonic and utilitarian aspects of consumer behavior: An attitudinal perspective. *Advances in Consumer research*, 12(1).

5. Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2), 179-211.

6. Alexa Skills Kit. developer.amazon.com. Policy Testing for an Alexa Skill, Advertising (5), from https://developer.amazon.com/it-IT/docs/alexa/custom-skills/policy-testing-for-an-alexa-skill.html#5-advertising.

7. Armentano, M. G., Abalde, R., Schiaffino, S., Amandi, A. (2014, November). User acceptance of recommender systems: Influence of the preference elicitation algorithm. In *2014 9th International Workshop on Semantic and Social Media Adaptation and Personalization* (pp. 72-76). IEEE.

8. Bentley, F., LuVogt, C., Silverman, M. Wirasinghe, R., White, B., Lottridge, D. (2018). Understanding the Long-Term Use of Smart Speaker Assistants. *Proc ACM Interact. Mod. Wearable Ubiquitous Technol*. 2, 3 Article 91 (September 2018).

9. Chaouali, W., Yahia, I. B., & Souiden, N. (2016). The interplay of counter-conformity motivation, social influence, and trust in customers' intention to adopt Internet banking

services: The case of an emerging country. *Journal of Retailing and Consumer Services*, 28, 209-218.

10. Childers, T. L., Carr, C. L., Peck, J., Carson, S. (2001). Hedonic and utilitarian motivations for online retail shopping behavior. *Journal of retailing*, 77(4), 511-535.

11. Choi, J., Lee, H. J., Sajjad, F., Lee, H. (2014). The influence of national culture on the attitude towards mobile recommender systems. *Technological Forecasting and Social Change*, 86, 65-79.

12. Chu, L. (2019). Why would I adopt a smart speaker?: Consumers' intention to adopt smart speakers in smart home environment (Master's thesis, University of Twente).

13. Cyr, D., Head, M., Ivanov, A. (2006). Design aesthetics leading to m-loyalty in mobile commerce. *Information & management*, 43(8), 950-963.

14. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

15. Doshi, S. (2019). Brief on Recommender Systems: Different types of recommendation methods used in industries. *Toward Data Science*. Retrieved February 10, 2019. From https://towardsdatascience.com/brief-on-recommender-systems-b86a1068a4dd.

16. Edu, J. S., Such, J. M., Suarez-Tangil, G. (2019). Smart Home Personal Assistants: A Security and Privacy Review. *arXiv preprint arXiv*:1903.05593.

17. Forgas, J. P. (1989). Mood effects on decision making strategies. Australian Journal of Psychology, 41(2), 197-214.

18. Gao, L., Bai, X. (2014). A unified perspective on the factors influencing consumer acceptance of internet of things technology. *Asia Pacific Journal of Marketing and Logistics*.

19. Ge, M., Delgado-Battenfeld, C., Jannach, D. (2010, September). Beyond accuracy: evaluating recommender systems by coverage and serendipity. *In Proceedings of the fourth ACM conference on Recommender systems* (pp. 257-260).

20. Gefen, D., Karahanna, E., Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *Management Information Systems Quarterly*, 27(1), 51–90.

21. George, D., Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: *Allyn & Bacon*.

22. Gliem, J. A., Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education*.

23. Gram-Hanssen, K., Darby, S. J. (2018). "Home is where the smart is"? Evaluating smart home research and approaches against the concept of home. *Energy Research & Social Science*, 37, 94-101.

24. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L. (1998). Multivariate data analysis (Vol. 5, No. 3, pp. 207-219). Upper Saddle River, NJ: *Prentice hall*.

25. Hassanein, K. S., Head, M. (2004, October). Building online trust through socially rich web interfaces. In *Proceedings of the 2nd annual conference on privacy, security and trust, Fredericton*, New Brunswick, Canada (pp. 15-22).

26. Hassanein, K., Head, M. (2007). Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. *International Journal of Human-Computer Studies*, 65, 689–708.

27. Herlocker, J. L., Konstan, J. A., Terveen, L. G., Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *Transactions on Information Systems* (TOIS), 22(1), 5-53.

28. Holt (2018). How Much Do Smart Speakers Really Hear and Who is Listening? *AI Pavillion Seminar*.

29. Hooper, D., Coughlan, J., Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic journal of business research methods*, 6(1), 53-60.

30. Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical reference services quarterly*, 37(1), 81-88. DOI: 10.1080/02763869.2018.1404391.

31. Hu, L. T., Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.

32. Hu, P., Lu, Y., Gong, Y. (2019). Speaking vs. Listening: Balance AI Conversation Attributes for Better Voice Marketing. Extended abstract, version: September 9, 2019.

33. Hwang, Y., Kim, D. J. (2007). Customer self-service systems: The effects of perceived Web quality with service contents on enjoyment, anxiety, and e-trust. *Decision support systems*, 43(3), 746-760.

34. Ingham, J., Cadieux, J. (2016, January). From E-shopping system quality to the consumer's intention to return: A meta-analytic study of the mediation of attitude, usefulness, enjoyment, and trust. In *2016 49th Hawaii International Conference on System Sciences* (HICSS) (pp. 3556-3564).

35. Jefferson, G. (1949). The mind of mechanical man. *British Medical Journal*, 1(4616), 1105.

36. Jung, H., Oh, C., Hwang, G., Oh, C. Y., Lee, J., Suh, B. (2019, May). Tell Me More: Understanding User Interaction of Smart Speaker News Powered by Conversational Search. In Extended Abstracts of the 2019 CHI *Conference on Human Factors in Computing Systems* (pp. 1-6).

37. Kennedy, John (2019). How AI Completed Schumbert's Unfinisched Symphony No 8. *siliconrepublic.com*. Retrieved February 22, 2019. From https://www.siliconrepublic.com/machines/unfinished-symphony-no-8-ai-huawei.

38. Kim, H. W., Chan, H. C., Gupta, S. (2007). Value-based adoption of mobile internet: an empirical investigation. Decision support systems, 43(1), 111-126.

39. Kim, Y. J., Chun, J. U., Song, J. (2009). Investigating the role of attitude in technology acceptance from an attitude strength perspective. *International Journal of Information Management*, 29(1), 67-77.

40. Kim, D., Park, K., Park, Y., Ju, J., & Ahn, J. H. (2018). Alexa, Tell Me More: The Effect of Advertisements on Memory Accuracy from Smart Speakers. In PACIS (p. 204).

41. Kinsella, B. (2019, February [1]). Google Assistant Actions Total 4,253 in January 2019, Up 2.5x in Past Year but 7.5% the Total Number Alexa Skills in U.S. *Voicebot.ai*. Retrieved February 15, 2019. From https://voicebot.ai/2019/02/15/google-assistant-actions-total-4253-in-january-2019-up-2-5x-in-past-year-but-7-5-the-total-number-alexa-skills-in-u-s/.

42. Kinsella, B. (2019, February [2]). Amazon Increases Global Smart Speaker Sales Share in Q4 2018, While Google's Rise Narrows the Gap and Apple Declines. *Voicebot.ai*. Retrieved February 20, 2019. From https://voicebot.ai/2019/02/20/amazon-increases-global-smart-speaker-sales-share-in-q4-2018-while-googles-rise-narrows-the-gap-and-apple-declines/.

43. Kinsella, B. (2019, March [3]). U.S. Smart Speaker Ownership Rises 40% in 2018 to 66.4 Million and Amazon Echo Maintains Market Share Lead Says New Report from Voicebot. *Voicebot.ai*. Retrieved March 7, 2019. From https://voicebot.ai/2019/03/07/u-s-smart-speaker-ownership-rises-40-in-2018-to-66-4-million-and-amazon-echo-maintains-market-share-lead-says-new-report-from-voicebot/.

44. Kinsella, B. (2019, October [4]). Amazon Alexa Has 100k Skills but Momentum Slows Globally. Here is the Breakdown by Country. *Voicebot.ai*. Retrieved October 1, 2019. From https://voicebot.ai/2019/10/01/amazon-alexa-has-100k-skills-but-momentum-slows-globally-here-is-the-breakdown-by-country/.

45. Kirakowski, J., Corbett, M. (1993). SUMI: The software usability measurement inventory. *British journal of educational technology*, 24(3), 210-214.

46. Koch, L. (2019). Smart Speakers Shopping Gains Traction. *eMarketer*. Retrieved July 8, 2019. From https://www.emarketer.com/content/smart-speaker-shopping-gains-traction.

47. Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security*, 64, 122-134.

48. Kowalczuk, P. (2018). Consumer acceptance of smart speakers: a mixed methods approach. *Journal of Research in Interactive Marketing* (pp. 418-431).

49. Lai, J. Y., Debbarma, S., Ulhas, K. R. (2012). An empirical study of consumer switching behaviour towards mobile shopping: a Push–Pull–Mooring model. *International Journal of Mobile Communications,* 10(4), 386-404.

50. Lee, M. (2009). Predicting and explaining the adoption of online trading: An empirical study in Taiwan. *Decision Support Systems*, 47, 133–142.

51. Marriott, H. R., & Williams, M. D. (2018). Exploring consumers perceived risk and trust for mobile shopping: A theoretical framework and empirical study. *Journal of Retailing and Consumer Services*, 42, 133-146.

52. McNee, S. M., Riedl, J., Konstan, J. A. (2006, April). Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems* (pp. 1097-1101).

53. Moon, J. W. Kim, Y. G. (2001). Extending the TAM for a WorldWide-Web context. *Information & Management*, vol. 38, pp. 217-230.

54. Nysveen, H., Pedersen, P. E., Thorbjørnsen, H. (2005). Intentions to use mobile services: antecedents and cross-service comparisons. *Journal of the Academy of Marketing Science*, 33(3), 330-346. DOI: 10.1177/0092070305276149.

55. Newman, N. (2018, November). The Future of Voice and the Implications for News. *Reuters Institute for the Study of Journalism*, University of Oxford. Digital News Project, retrieved: November 2018.

56. Oh, Jasmine (2019). Yes, AI Will Replace Designers: But Here Are 3 Ways We Can Work With AI For Future Success. *medium.com*. Retrieved March 28, 2019. From https://medium.com/microsoft-design/yes-ai-will-replace-designers-9d90c6e34502.

57. Park, Y., Son, H., Kim, C. (2012). Investigating the determinants of construction professionals' acceptance of web-based training: an extension of the technology acceptance model. *Automation in Construction*, 22 (2012), pp. 377-386.

58. Pu, P., Chen, L., Hu, R. (2011). A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems* (pp. 157-164).

59. Qiu, L., Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of management information systems*, 25(4), 145-182, DOI: 10.2753/MIS0742-1222250405.

60. Raghunathan, R., Corfman, K. (2006). Is happiness shared doubled and sadness shared halved? Social influence on enjoyment of hedonic experiences. *Journal of Marketing Research*, 43(3), 386-394.

61. Ricci, F., Rokach, L, Shapira, B., Kantor, P. B. (2011). Recommender Systems Handbook. *Springer*. Boston, MA. DOI: 10.1007/978-0-387-85820-3.

62. Robinson Jr, L., Marshall, G. W., Stamps, M. B. (2005). Sales force use of technology: antecedents to technology acceptance. *Journal of Business Research*, 58(12), 1623-1631.

63. Rodrìguez, G. (2019). Introduction to Recommender Systems in 2019. *Tryo Labs*. Retrieved 2019. From https://tryolabs.com/blog/introduction-to-recommender-systems/.

64. Rogers, E. M. (2010). Diffusion of innovations. *Simon and Schuster*.

65. Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*, 48(2), 1-36.

66. Sarkar, A. (2011). Impact of utilitarian and hedonic shopping values on individual's perceived benefits and risks in online shopping. *International management review*, 7(1), 58-65.

67. Sciuto, A., Saini, A., Forlizzi, J., Hong, J. I. (2018, June). "Hey Alexa, What's Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 857-868).

68. Shardanand, U. (1994). Social information filtering for music recommendation (Doctoral dissertation, Massachusetts Institute of Technology).

69. Smart-home worldwide, Statista. https://www.statista.com/outlook/283/100/smart-home/worldwide.

70. Smith, K. T. (2018). Marketing via Smart Speakers: What Should Alexa Say?. *Journal of Strategic Marketing*, 1-16.

71. Sohn, K., Kwon, O. (2020). Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products. *Telematics and Informatics*, 47, 101324.

72. Suh, B., Han, I. (2002). Effect of trust on customer acceptance of Internet banking. *Electronic Commerce research and applications*, 1(3-4), 247-263.

73. Swanson, E. B. (1987). Information Channel Disposition and Use. *Decision Sciences*, vol. 18, pp. 131-145.

74. Taylor, S., Todd, P. (1995). Assessing IT usage: The role of prior experience. *MIS quarterly*, 561-570.

75. *TechTarget* (2018, January). Voice Recognition (Speaker Recognition). *TechTarget*. Retrieved January 2018. From: https://searchcustomerexperience.techtarget.com/definition/voice-recognition-speaker-recognition.

76. Tintarev, N., Masthoff, J. (2007, April). A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop* (pp. 801-810).

77. Tung, F., Chang, S., Chou, C. (2008). An extension of trust and TAM model with IDT in the adoption of the electronic logistics information system in HIS in the medical industry. *International Journal of Medical Informatics*, 77, 324–335.

78. Turing, A. M. (2009). Computing machinery and intelligence. In Parsing the Turing Test (pp. 23-65). *Springer*, Dordrecht.

79. Underwood, C. (2020). Use Cases of Recommendation Systems in Business – Current Applications and Methods. *Emarje, the AI and Research and Advisory Company*. Retrieved March 4, 2020. From https://emerj.com/ai-sector-overviews/use-cases-recommendation-systems/.

80. Venkatesh, V., Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2), 186-204.

81. Venkatesh, V., Morris, M. G., Davis, G., Davis, F. D. (2003). User acceptance of information technology: toward a unified view. *Mis Quarterly*, 27(3), 425-478.

82. *Voicebot* (2019, January). Consumer Smart Speaker Consumer Adoption Report. *Voicebot.ai*.

83. Wang, J., De Vries, A. P., Reinders, M. J. (2006, August). Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 501-508).

84. Wang, Y. Y., Townsend, A., Luse, A., Mennecke, B. (2012). The determinants of acceptance of recommender systems: Applying the UTAUT model. *AMCIS 2012 Proceedings*. 2.

85. Warshaw, P. R. (1980). A new model for predicting behavioral intentions: An alternative to Fishbein. *Journal of marketing research*, 17(2), 153-172.

86. Wharton, S. (2018). Many Consumers Still Concerned About Smart-Home Devices. *eMarketer*. Retrieved March 13, 2018. From https://www.emarketer.com/content/tbd.

87. Yang, H. D., & Yoo, Y. (2004). It's all about attitude: revisiting the technology acceptance model. *Decision support systems*, 38(1), 19-31.

88. Zhao, J., Fang, S., Jin, P. (2018). Modeling and quantifying user acceptance of personalized business modes based on TAM, trust and attitude. *Sustainability*, 10(2), 356.

89. Ziegler, C. N., McNee, S. M., Konstan, J. A., Lausen, G. (2005, May). Improving recommendation lists through topic diversification. *In Proceedings of the 14th international conference on World Wide Web* (pp. 22-32).

APPENDIX

Here, the table exhibits the questions provided in the survey questionnaire. The scale items (questions range from PEoU_1 to AC_3) were collected in 5-point likert scales (1 – completely disagree; 5 – completely agree), they are adaptations of well-established scales from previous research studies.

CONSTRUCT	QUESTION ITEM	SCALE	REFERENCE
	Do you know smart speakers (devices like Amazon Alexa and Google Home)?	Yes/No	
	Have you ever used a smart speaker?	Yes/No	
	Do you own a smart speaker?	Yes/No	
	How long have you owned a smart speaker?	Less than one month; from one to six months; from six months to one year; from one to two years; more than two years.	
Perceived Ease of Use	PEoU1: Using the smart speaker would be easy.	1 – completely disagree; 5 – completely agree.	Venkatesh, V., & Davis, F. D. (2000)
	PEoU2: Interacting with the smart speaker does not require a lot of my mental effort.	1 – completely disagree; 5 – completely agree.	
	PEoU3: I would find the smart speaker difficult to use.	1 – completely disagree; 5 – completely agree.	Sohn, K., & Kwon, O. (2020)
Perceived Usefulness	PU1: If my smart speaker offered to me a song to listen to, it would help me.	1 – completely disagree; 5 – completely agree.	Sohn, K., & Kwon, O. (2020)
	PU2: I would find useful that my smart speaker offered me a song to listen to.	1 – completely disagree; 5 – completely agree.	Venkatesh, V., & Davis, F. D. (2000)
	PU3: If my smart speaker offered me a song to listen to, it would improve the efficiency of my music research.	1 – completely disagree; 5 – completely agree.	
Enjoyment	E1: I would have fun interacting with my smart speakers if it offered me a song to listen to.	1 – completely disagree; 5 – completely agree.	Agarwal, R., & Karahanna, E. (2000)
	E2: If my smart speakers offered me a song to listen to, it would bore me.	1 – completely disagree; 5 – completely agree.	
	E3: The suggestion of listening to a song from my smart speaker should provide me with a lot of enjoyment.	1 – completely disagree; 5 – completely agree.	
Social Influence	SI1: People who are important to me think that I should listen to the songs offered by my smart speaker.	1 – completely disagree; 5 – completely agree.	Venkatesh, V., & Davis, F. D. (2000)
	SI2: People who influence my behaviour think that I should listen to the songs offered by my smart speaker.	1 – completely disagree; 5 – completely agree.	
	SI3: People around me would think that I should not listen to the songs offered by my smart speaker.	1 – completely disagree; 5 – completely agree.	Sohn, K., & Kwon, O. (2020)

Trust	T1: I believe that my smart speaker is honest when it offered me a song to listen to.	1 – completely disagree; 5 – completely agree.	Chu, L. (2019)
	T2: I believe that my smart speaker is trustworthy when it offered me a song to listen to.	1 – completely disagree; 5 – completely agree.	
	T3: If my smart speaker offered me a song to listen to, it would be on my best interest.	1 – completely disagree; 5 – completely agree.	Qiu, L., & Benbasat, I. (2009)
Attitude	AT1: I like the idea that my smart speaker offers me a song to listen to.	1 – completely disagree; 5 – completely agree.	Chu, L. (2019)
	AT2: I would have negative feelings toward my smart speaker if it offered me a song to listen to.	1 – completely disagree; 5 – completely agree.	
	AT3: It would be good for my daily life if my smart speaker offered me a song to listen to.	1 – completely disagree; 5 – completely agree.	Sohn, K., & Kwon, O. (2020)
Acceptance	AC1: I intent to listen to a song offered by my smart speaker in the future.	1 – completely disagree; 5 – completely agree.	Sohn, K., & Kwon, O. (2020)
	AC2: I am willing to listen to a song offered by my smart speaker in the near future.	1 – completely disagree; 5 – completely agree.	Chu, L. (2019)
	AC3: I will not listen to a song offered by my smart speaker in the near future.	1 – completely disagree; 5 – completely agree.	
Demographics	Gender:	Male/Female	
	Age:		
	Employment:	Student; process worker; office worker; freelance; unemployed; other	

If the respondent answers “No” to the first question (Do you know smart speakers [devices like Amazon Alexa and Google Home]?), The following brief description of smart speakers was provided, together to the two images of Amazon Echo and Google Home:

> *“Smart speakers are table devices connected to the Internet, which are able to communicate with other connected devices (such as smartphones and others). The most famous ones are: Amazon Echo (“Alexa”) and Google Home. These devices are activated with a wake word (e.g. “Hey Google!” or “Alexa!”) and by voice it is possible to impose some commands, such as: listen to music; turn on and off the light; ask for weather information; set alarms and chronometers; surfing in the Internet…”*

The picture of Amazon Echo is the following:

Whereas, the picture of Google Home is shown in the figure below:

ABSTRACT

Research gap and research questions

Once the smart speakers have been recognized by many marketers as a new and fast developing marketing channel, studying users' behaviours and adoption mechanisms becomes crucial to improve their business and extract more value from both potential and current customers. Therefore, many studies concerning users' acceptance of smart speakers were conducted in the last years.

Since many efforts to understand the users' acceptance of smart speakers have already been spent, this work is focus on the adoption mechanisms and antecedents of one of the major opportunities provided by these devices, I mean, the recommendations that the devices can offer to the users. Recommendations are an effective tool to enhance user experience and cope with the information overload, which affects users' decision making on most of the online environments. Recommendations can be provided by smart speakers through software tools called Recommender Systems (RS). A wide range of researches on recommender systems attempts to understand which algorithm achieve the best performances, while the best performances themselves have been identified along many dimensions (accuracy, variety, etc.); while other studies deal with the acceptance of recommender systems in several circumstances.

However, there are not studies concerning the user acceptance of recommender systems from smart speakers. The question is critical since many companies have adopted a successful business model based on recommender systems (Amazon, Netflix, etc.). In this stage of market expansion, a proper understanding of users' acceptance can become a crucial competitive advantage for the companies which hope to extract customers' value equipping their apps for smart speaker with a voice recommender system. In order to achieve a comprehensive understanding of user acceptance of such software, many variables and their relationships must be studied.

This study aims to fill up this gap and to provide a comprehensive insight of users' acceptance of such systems from smart speakers, underlying the main variables that affect the acceptance and the relationships among them. The deeper understanding of the variables that affect the acceptance of recommender systems from smart speakers and the acceptance model proposed on this study represent the main theoretical contributions of this work, since no previous researches involved the study of acceptance of RSs from smart speakers. Then, practical contributions are provided with the suggestions, inferred by the results of the data analysis, for the marketers who want to improve their business developing recommendation systems for smart speakers.

Thus, the research questions of my study can be formulated as follows:

- *Which are the main variables that affect user acceptance of recommender systems of smart speakers*?
- *Which are their relationships*?
- *How can firms behave in order to receive the best response in term of acceptance of such software equipped to smart speakers*?

Theoretical framework and hypotheses

Starting from original TAM (Technology Acceptance Model) by Davis, we developed a user's acceptance model of recommender systems from smart speakers. This model has been chosen because it was corroborated by many studies and it is considered a robust acceptance model. Moreover, it was adopted in studies concerning both recommender systems and smart speakers.

The TAM accounts for three antecedents of user's acceptance: Perceived Ease of Use, Perceived Usefulness and Attitude. To enrich the model, other variables are been added to the TAM in order to obtain a comprehensive framework. Enjoyment, Trust and Social Influence are considered important factors that affect user acceptance of recommender systems from smart speakers. Thus, 7 variables are introduced in the model:

1) **Perceived Ease of Use** (**PEoU**): the degree to which a person believes that using a particular system is free of effort.
2) **Perceived Usefulness** (**PU**): the degree to which a person believes a particular system would enhance his or her performance.
3) **Attitude** (**AT**): comprehensive judgement or opinion toward the recommender system.
4) **Social Influence** (**SI**): the degree to which an individual perceives that important others believe he or she should use the recommender system.
5) **Enjoyment** (**E**): the extent to which the activity of using a specific system is perceived to be enjoyable in its own right, aside from any performance consequences resulting from system use.
6) **Trust** (**T**): the degree of which the recommender systems from smart speakers are found trustworthy and they are in users' best interests.
7) **Acceptance** (**AC**): the acceptance of recommender systems from smart speakers concerns the intention to use the systems.

The resulting model is exhibited below:

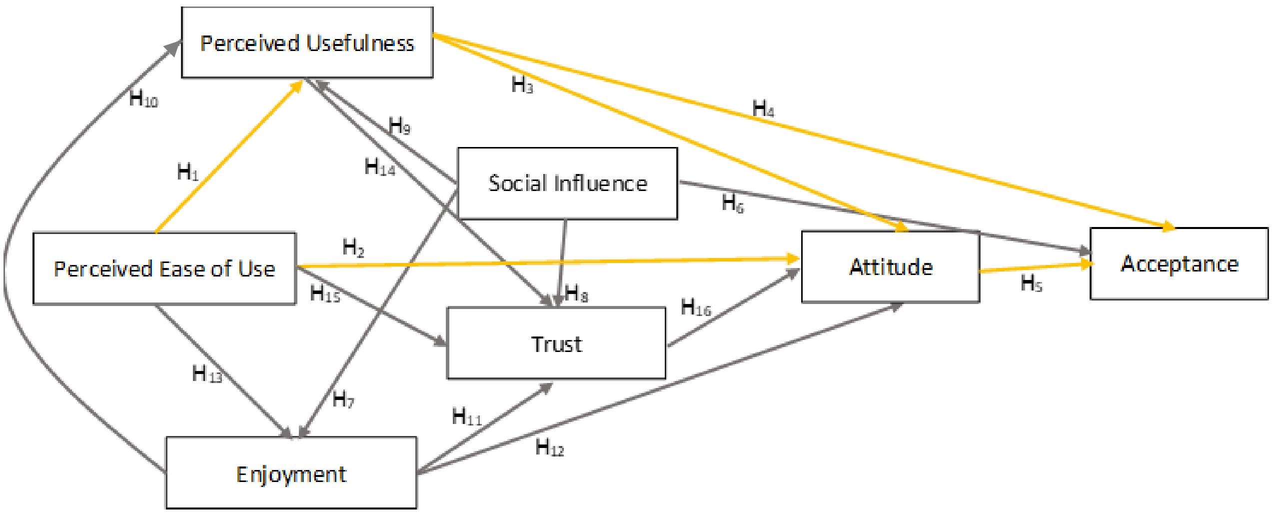

Figure 1 (Abstract): The figure shows the hypothesized relationships between the variables analysed in the model. Each arrow represents a hypothesis. The yellow arrows represent the original TAM by Davis.

Research design and methodology

In order to answer the research questions, a survey questionnaire was developed (the questions are available in the Appendix to this work). The scale variables were gathered in 5-point likert scales (1 – completely disagree; 5 – completely agree), then some demographics were asked (gender, age and employment). The questions refer to the offering to listen to a song made by the recommender system from a smart speaker. The music was chosen because both males and females feel comfortable with it and there is corroborated evidence that the smart speakers are mostly used to listen to music.

The questionnaire was undertaken to my contacts and posted in university social networks to reach as many respondents as possible, during the period between 4th March and 13th March 2020. In total, 286 responses were collected. The average age of the respondents is 23.96 years (standard deviation 5.69, 76 missing values), while the sample is made up 141 females (67%, 75 missing values). The 74% (212) of the respondents completed the entire survey, one of those missed only the demographic box. The respondents are mainly students (84%, 76 missing values). Only the 212 respondents who completed the scales questions were included in the data analysis.

Data analysis

Here, Structural Equation Modeling (SEM) was performed to test the hypotheses. The open source R environment was employed, using the "lavaan" package from Rosseel (2012). It performs an iterative method, involving the definition of a set of *start values* of the free parameters. These default values are corrected at each iteration comparing the resulting covariance matrix with the observed (actual) covariance matrix. The iterations end when no change occurs (or a tolerance difference is reached).

The SEM supports 7 out of 16 hypotheses. The supported hypotheses are shown in the table below.

N.	HYPOTHESIS	β-COEFFICIENT	P-VALUE	EVALUATION
H_4	PU → AC	0.63	< 0.01	supported
H_5	AT → AC	0.14	< 0.05	poorly supported
H_7	SI → E	0.38	< 0.01	supported
H_{10}	E → PU	1.00	< 0.01	supported
H_{11}	E → T	0.74	< 0.01	supported
H_{12}	E → AT	0.78	< 0.01	supported
H_{16}	T → AT	0.21	< 0.01	supported

Table 1 (Abstract): Final model hypotheses evaluation, β coefficients

Then, the final users' acceptance model for recommender systems from smart speakers is schematized in the following diagram:

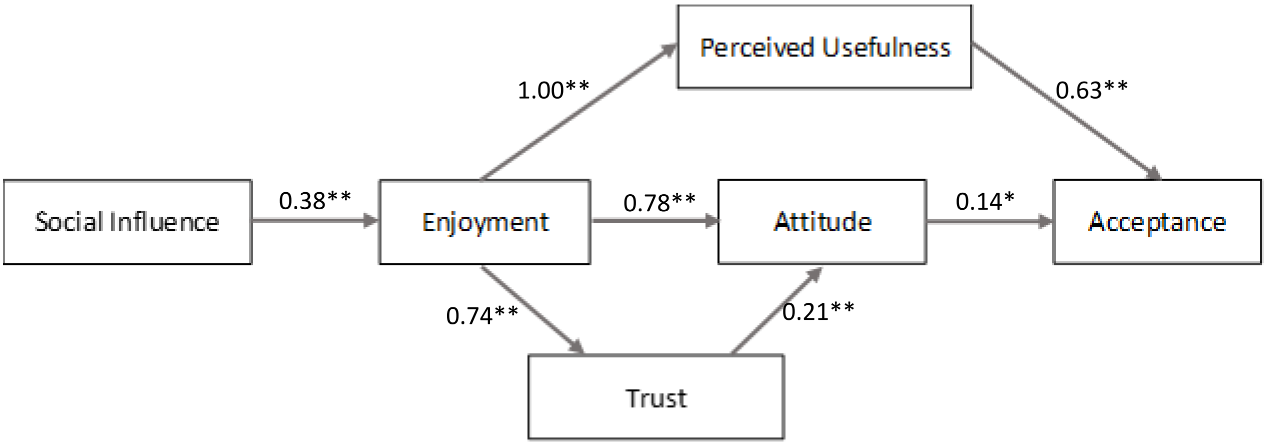

*Figure 2 (Abstract): Final model of user acceptance of recommender systems from smart speakers. β coefficients are represented on the arrows. * p-value<0.05; ** p-value<0.01.*

Discussion

The weights of each variable on the explanation of user's acceptance of recommender systems from smart speakers are exhibited in the following chart:

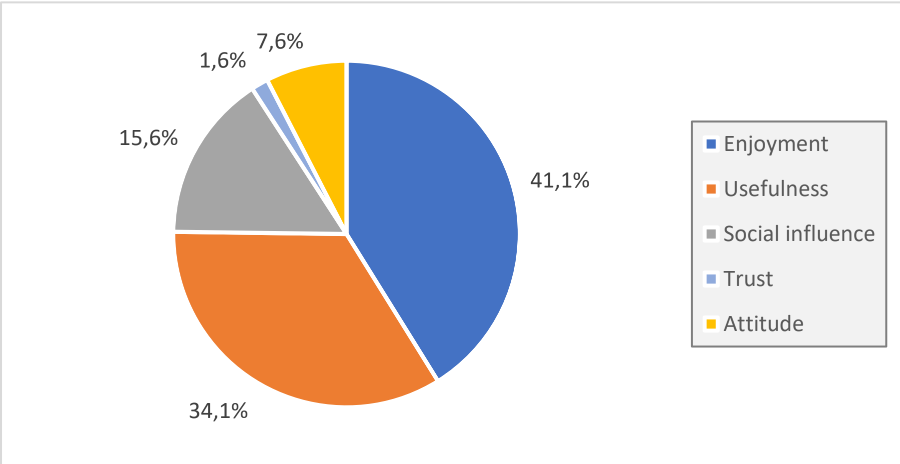

Figure 3 (Abstract): The proportion of the user's acceptance of recommender systems from smart speakers explained by each of the variable analysed in the final model.

The crucial role of **Enjoyment** emerges from a plethora of previous researches on usage and acceptance of smart speakers, and it was confirmed even on acceptance of recommendation systems from such devices, where the E is the most prominent variable that explain the Acceptance.

Another well-corroborated driver of acceptance is the **Perceived Usefulness** (PU) of smart speakers, in general, and of RS for such devices, in particular. This is the second most important variable on the explanation of user's acceptance of RS from smart speakers.

The **Perceived Ease of Use**, which is a salient variable in TAM, does not exhibit any significant effect on other variables of the model. The place of PEoU is taken by the Enjoyment, which is also the most effective driver of acceptance of recommender systems from smart speakers. It is meaningful that E strongly affects both the PU and the Attitude, as PEoU makes in the Davis' TAM. This means that the Perceived Ease of Use does not have a prominent role on explaining the acceptance in this peculiar case. The relatively poor effort which is required to the user of smart speakers in the usage phase can explain the marginal role of PEoU. Voice commands are easy to deliver and both the people who experimented the smart speakers and the ones who have never used them perceive this ease of use. It is not by chance that PEoU scale scored the highest values (mean = 3.85 out of 5.00) among the constructs.

By contrast, the **Social Influence** poorly affects the E, while neither no significant effect on Trust and PU nor a direct effect on acceptance were discovered. The SI exhibits also the lowest average scores, meaning that perceived social approval is not a salient issue on the acceptance of recommender systems from smart speakers.

A marginal role was relegated also to **Trust**, which mediates the E and the AT, but its effect on Acceptance is very poor. Although security and privacy concerns affect the smart speakers, they do not result in a negative perception of the trustworthiness of the recommender systems from such devices. Many studies encourage this interpretation, since people claim to be sensitive to privacy and security issues, but in their actual behaviours these concerns have poor, or even not at all, relevance. This inconsistency of privacy attitudes and privacy behaviour is often referred to as the “privacy paradox”.

Suggestions

Enjoyment is a crucial variable that the marketers can exploited when advertise from smart speakers or when they built recommender systems for such devices. Funny and playful ways to recommend items seem to strongly affect the acceptance, and then, the adoption and the persuasiveness of RS in this field.

Recommendation systems are known to be an effective way to cope with the information overload issue. The marketers should leverage on the possibility to have more personalized results when looking for offerings to purchase and on the effectiveness of the recommendations which helps the user to save time and mental effort on their researches. Actually, the researches could be time-consuming and the unsatisfied consumers are annoyed and frustrated when they do not achieve their purposes. Even the user’s attention decreases when the time spent on the research rises. Therefore, the marketers must teach to the users that the functionalities of the RS are on users’ best interests and they could be useful to enhance user’s efficacy and effectiveness of the research.

Moreover, the performance of the RSs must be evaluated both by accuracy and serendipity measures. The accuracy reflects the functional aspect of recommender systems and it concerns the degree by which the recommended items reflects the actual users’ preferences, then it measures the usefulness of the suggestions. Whereas, the serendipity measures the users’ emotional responses, then it is more suitable to evaluate the playfulness and fun of the RS and the respective recommended items.

Hence little or no differences are been found between males and females, it is not suggested to marketers to leverage on this group distinction. Contrarily, the respondents who have used the smart speakers at least once show a significant higher acceptance than the respondents who have never used such devices. This means that the acceptance of recommender systems could be improved just encouraging the trial of such systems, perhaps by decreasing the costs associated with the systems (ease of installation of the “skills”, decreasing the price of smart speakers) or by offering promotions to the customers who use the RS from smart speakers.

Limitations and future researches

The research study we have conducted has some limitations that we need to underline. The first one is the poor variability of the sample of respondents who participated the experiment. Almost all the respondents are students with an age that ranges from 19 to 27. Further researches could be performed in order to discover whether the age or the social status of the users significantly affects the acceptance model we have proposed. Even the nationality could have a role on shaping the weights of the variables, while the proposed study was conducted only on Italian respondents.

The recommendation item we considered was the songs, since the music applications are the most used from smart speakers. Nevertheless, marketers are also interested on selling products, then further researches could focus on users’ purchase intentions elicited by RS from smart speakers.

Marketers can build their own third-party skills to promote their offerings. The marketers could equip these applications with recommender systems in order to take advantage from this technology. Further researches would involve the role played by the trustworthiness of the third-party skill provider on shaping the users’ willingness to adopt and accept the RSs.

Another limitation of this work is represented by the Attitude (AT) scale, which is made of only one question item. Attitude is a complex construct which needs a multi-item scale to yield more reliable results. Even though, in order to maintain internal reliability and external validity of the constructs, we decided to reduce this construct to a single-item scale. Further researches could leverage on this disadvantage and build more sophisticated measures of AT.

Moreover, it would be considered to enrich the Attitude measure distinguishing the cognitive from the affective attitude. The cognitive attitude deals with the believes about the technology, while the affective one is about favourability and likeness. Yang & Yoo (2004) tested the TAM along the cognitive and affective dimensions of attitude. They administered a survey questionnaire to U.S. students and found that cognitive dimension has a significant stronger mediator effect on Acceptance than the affective one. The quoted study would suggest to looking for more complex constructs of attitude in order to reach more reliable results.