Asia Pacific Journal of Information Systems
Vol. 33 No. 4 (December 2023), 999-1015

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2023.33.4.999

Privacy Concerns of Smart Speaker Users in South Korea: A Text-mining Analysis

Hong Joo Lee[a], Guglielmo Maccario[b,*], Maurizio Naldi[c]

[a] *Professor, Department of Business Administration, The Catholic University, Korea*
[b] *Contract Professor, UNINT University of the International Studies, Italy*
[c] *Professor, GEPLI Department, LUMSA University, Italy*

ABSTRACT

Smart speakers represent a growing product in home electronics. However, their capability to record voices in their immediate surroundings has spurred concerns about privacy violations. In this paper, we assess the extent of those concerns in the opinions of smart speaker users by examining the reviews posted by smart speaker users. We focus on South Korea as a representative of advanced Asian economies. The results show that Korean smart speaker users are either unconcerned or unaware of privacy issues, confirming the results of previous studies about UK users, but with an even lower degree of interest in the topic. However, for the few users concerned about privacy, their attitude towards privacy influences their overall opinion about smart speakers.

Keywords: Smart Speakers, Privacy, South Korea, Sentiment Analysis

Ⅰ. Introduction

While Google, Amazon, and Apple are under pressure to enable their smart speakers to collect and store voice recordings of their users, a similar situation is unfolding in South Korea. In September 2019, the Korean tech giants Naver and Kakao admitted that they had been gathering audio data from smart speaker users and converting it to written files. Furthermore, identical methods were also discovered to have been carried out by KT and SK Telecom to improve the performance of their devices (Yeo, 2019). Potential privacy violations in smart speakers have hit the news and have become a hot issue.

Speakers equipped with intelligent voice assistants, like Google Assistant and Amazon Alexa, are used every day for entertainment, question-answering, web browsing, and a variety of other tasks. These devices are the core control points for many smart home features, and their popularity has increased

*Corresponding Author. E-mail: guglielmomc@gmail.com

enormously in the last eight years. Since the launch of the first Amazon Echo in 2014, the use of smart speakers has spread internationally, with global shipments expected to top 200 million in 2023.[1] Smart speakers have been adopted at a global level, and the privacy issues related to their use have attracted the attention of academia as well as the media. Since the very first stage of smart speakers' adoption, practitioners and academicians have shown an interest in such devices' privacy and security vulnerabilities. Despite the field's novelty, scholars have created a substantial corpus of relevant literature, exploring different themes while discussing how using smart speakers might expose users to privacy and security risks. Previous research in the field of information privacy of smart speakers has been exploring the perception of privacy issues (Abdi et al., 2019; Bolton et al., 2021), their influence on adoption decisions (Bawack et al., 2021; Pridmore and Mols, 2020), the identification of vulnerabilities (Jiang and Apthorpe, 2021; Zhang et al., 2019), the proposal of possible countermeasures (Lin et al., 2021; Mandalari et al., 2020) and the study of legal implications (Neville, 2020). An exhaustive literature review of the privacy risks of smart speakers is proposed by Maccario and Naldi (2022b) and Bolton et al. (2021).

However, in spite of the many countries affected by the phenomenon, the field of privacy of smart speakers has been largely limited to certain themes and national boundaries, lacking a cross-cultural, multinational approach. Nearly every nation has systems for regulating privacy, yet each one has its own distinctive set of behavioural and psychological approaches (Knijnenburg et al., 2022), reflecting the national character and culture. In their multinational study on online privacy, Cho et al. (2009) discovered that national culture strongly influenced internet users' privacy concerns. Similarly, Bellman et al. (2004) found that consumers in different countries live in diverse social environments, which may result in considerable disparities in levels of concern between nationalities. Recently, Maccario and Naldi (2022a) reported the findings of a text-mining analysis of over 4,500 reviews of the Echo line of smart speakers collected from Amazon, which found the owners to be generally unconcerned about privacy issues.

However, their conclusion was based on a sample of reviews limited to a specific Western country: the United Kingdom. How do customers from culturally and geographically diverse backgrounds perceive and express privacy concerns about their smart speakers? Would a text-mining analysis based on materials collected from a culturally diverse country reach results similar to what was obtained for the United Kingdom? In particular, would that occur when we consider a country from a very different geographical area, e.g., a different continent, embodying a different culture? Since the United Kingdom is a significant representative of Western Europe, Asia could represent an alternative setting to explore users' attitudes. In particular, we could examine a country with a similar level of technological development and consumption but a different national cultural background. Our choice has been South Korea. South Korea and the UK are known for their cultural differences (Hofstede, 2003; Nisbett, 2003), exhibiting disparities in collectivism and uncertainty avoidance (Choi et al., 2014), which in turn influence distinct behaviors of adopting IT services.

In this paper, we examine the attitude of smart speakers owners towards privacy in South Korea. We replicate the methodology employed in Maccario

1) Source: Global smart device shipment forecasts 2020 to 2023, https://www.canalys.com/newsroom/canalys-worldwide-smart-device-shipments-2023

and Naldi (2022a) for the sake of performing a comparison between a Western country and an Asian one on level ground. The research questions we wish to answer are:

RQ1: Is privacy a significant topic for smart speaker buyers in South Korea?
RQ2: What is the sentiment of smart speaker buyers about privacy in South Korea?
RQ3: Does the sentiment toward the privacy component influence the overall sentiment about smart speakers in South Korea?

The answers to those questions allow us to draw a parallel between the attitude of Western and Asian consumers. As in Maccario and Naldi (2022a), we rely on the opinions expressed by consumers on e-commerce platforms.

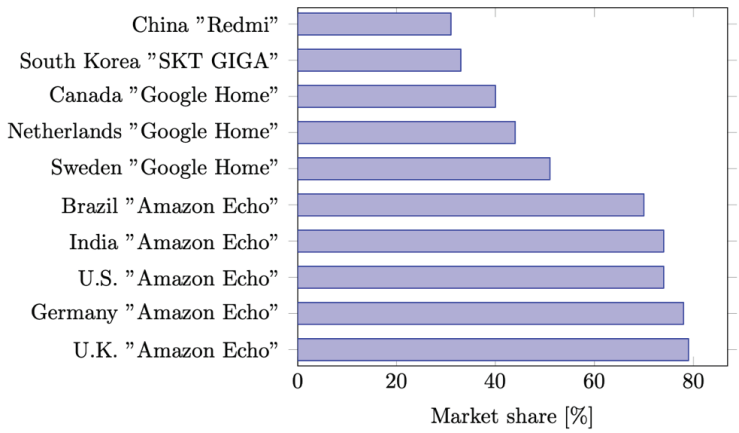
In Section 2, we review the market of smart speakers in South Korea. In Section 3, we describe the privacy features incorporated in the most popular smart speakers in the Korean market. In Section 4, we detail the rationales behind selecting platforms in South Korea for gathering reviews and blog posts. Methods and results are described in Sections 5 and 6, respectively.

Ⅱ. The Market of Smart Speakers in South Korea

The market of smart speakers in South Korea is markedly different from that of Western countries, which was analysed in Maccario and Naldi (2022a). In this section, we describe its most relevant characteristics. The increasing importance of smart speakers within the private houses of South Koreans is confirmed by the latest Smart Home Report by Statista, which reveals how the number of South Korean households actively using a smart home appliance is expected to reach 13.8m users by 2025; with a household penetration of 20.6% in 2021, expected to reach 61.5% by 2025.[2] Despite a growth rate similar to other western countries, the South Korean market of smart speakers is unique in terms of competitors and composition. As shown in <Figure 1>, Amazon and Google have nearly no direct competitors in all the countries surveyed in the Global Consumer Survey 2021 by Statista, whereas China and South Korea are the only exceptions.

The Korean smart speaker landscape is characterized by several national companies competing with the most prominent American brands. Amazon, Google, and Apple are battling for dominance with South Korean companies such as SKT, KT, and Kakao, and it seems like they are losing ground. According to the latest Statista Global Consumer Survey, as of October 2022, the two most owned smart speakers among respondents were developed by Korean companies: KT and SKT, with a market share respectively of 24% and 17%, followed by Google Home, owned by 16% of the respondents, Naver Wave (11%), and Kakao mini (11%). As it can be seen in <Figure 2>, Amazon Echo was reported to be owned by only 3% of the respondents. It can be assumed that, similarly to what happened in China, one of the key reasons that allowed South Koreans manufacturers to be competitive was the rapidity in launching their own product shortly after the debut of the first generation of Amazon Echo in the United States. The first South Korean company to unveil its smart speaker was SKT with its smart speaker Nugu, followed by KT with its GiGa Genie, Naver

2) https://www.statista.com/forecasts/887726/smart-home-penetration-rate-per-segment-in-south-korea



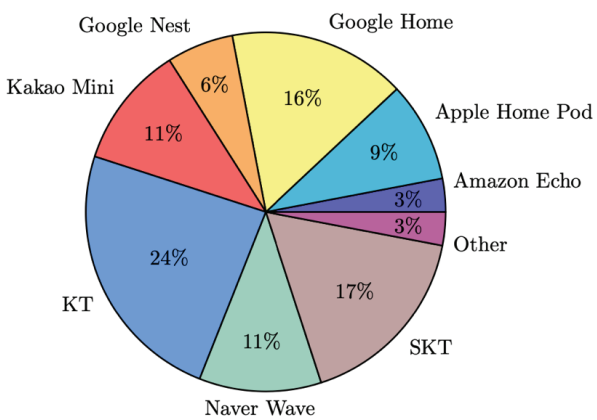
<Figure 1> Most Owned Smart Speaker Brands in 2021, by Country

with Clova and Kakao with Kakao mini (Malin, 2019).

For the purpose of our research, we extracted and analysed reviews and blog posts concerning five products sold on the Korean smart speakers' market. The five devices that were selected for our analysis are:

- SKT NUGU Candle;
- SKT NUGU Mini;
- Kakao Mini C;
- Amazon Echo dot;
- Google Home Mini.

As observed in our brief market analysis in the paragraph above, the five selected devices offer a representative portrait of the current South Korean smart speaker landscape, jointly accounting for almost half of the smart speakers owned by South Koreans (exactly 47%). In order to capture the peculiarities of the South Korean market, we selected two of the most famous national brands (SKT and Kakao) and two U.S. companies (Amazon and Google).



<Figure 2> Smart Speaker Ownership by Brand in South Korea.
(Source: Statista Global Consumer Survey)

Ⅲ. Privacy Settings

Scholars have found that smart speaker customers may not know or even ignore which data practices and controls smart speaker companies have introduced on their devices (Ackerman et al., 1999; Clark et al., 2015; Tabassum et al., 2019).

Yet, many devices include privacy options that can be used to secure consumer data. In this section, we review the privacy settings of the five smart speakers we have selected.

3.1. SKT Smart Speakers: Candle and Mini

Each SKT smart speaker come with two dedicated online manuals: a product manual and a service manual. In the product manual, the only setting described which could be beneficial for users to protect their data is the Microphone OFF button, a feature that allows disabling voice recognition. A red LED shows that the microphones have been turned off, and the device will not respond to the wake word.

Furthermore, in the service manual, under the section "Hidden features of Nugu",[3] there are instructions on how to prevent the device from hearing sound while watching tv - a useful feature, since scholars have reported that misactivations due to background voices, like those produced by televisions, represent a major source of accidental triggers (Dubois et al., 2020).

3.2. Kakao

The informational webpage dedicated to Kako Mini smart speaker does not directly explain which features users can adopt to protect their privacy. The QA pages for Hey Kakao, the operating system of Kakao Mini C, provide information on privacy settings. Similar to other speakers, it features a wake-word ("Hey, Kakao") and a microphone-off button. In addition, several pages with extensive information on Kakao privacy policies can be found at the bottom of the page.[4] Kakao even developed an "easy-to-understand privacy policy" page with simplified information regarding their privacy policies on matters such as collection, use and deletion of personal data.[5]

3.3. Amazon Echo dot

Following reported cases and research on misactivations voice-recognition flaws of its smart speakers (Maccario and Naldi, 2022a), Amazon developed several web pages to explain how to manage its devices' privacy settings.[6]

Like other smart speakers in the Echo line, the Echo Dot is designed to record and send any voice interaction just after hearing a wake word, which is by default "Alexa" for all the Echo devices but can be manually changed to alternative wake words. In addition, Amazon Echo devices allow their users to create voice profiles, used to train Alexa, and assign the trained voice to the user account (Sudharsan et al., 2019).

Similar to what is found in SKT devices, the Echo Dot allows their users to disconnect the microphones of their device by clicking on the microphone-off button, which comes with a red LED indicating that the device will not respond until switched on again.

3) Please note that we here refer to the English translation of a South Korean webpage

4) https://www.kakao.com/policy/privacy

5) https://www.kakao.com/policy/privacyPolicy/easy?type=easy&lang=

6) https://www.amazon.com/gp/help/customer/display

Furthermore, Amazon claims that all users' recordings in transit to their cloud service are encrypted to prevent any fraudulent access.[7)]

3.4. Google Home Mini

Google Home users have options equivalent to those provided by Amazon to protect the privacy of their data. Users can always go to the Assistant Activity page to see what Google Home has recorded, set up automatic data deletion, or use voice commands to ask Google Assistant to delete the data. "Hey Google, remove what I said (this week, today, etc.)" would be the command in English. Furthermore, the Google Home Mini micro-phones, like all other devices examined, can be turned off by clicking on the turn-off button. Google also created an extensive webpage containing information on the privacy policies of their home products.[8)]

Each speaker features options to disable the microphone and necessitates wake words for initiating recordings. SKT NUGU includes a voice record button, allowing users to engage in conversation with the speaker without uttering the wake word. Conversely, all speakers except SKT NUGU present a dedicated page for users to opt out of utilizing voice recordings for service improvement. Amazon Echo, Google Home, and Kakao Mini also enable users to erase voice recordings and select the data retention period. However, this functionality is not explicitly outlined in the service manual or product pages for speakers from SKT.

Ⅳ. Selection of Platforms

According to the latest figures by the World Bank, South Korea is leading the top 10 list for internet penetration in Asia, right after Brunei (which has a population of 441,532).[9)] The internet penetration rate in South Korea is 96.50%. Thus, it is not unexpected that online buying has flourished in the nation. In South Korea, the e-commerce websites with the highest reach in 2022 were Coupang.com and 11Street,[10)] while the website with the overall highest reach was found to be Naver.com (as of March 2022, around 82.2% of computer users in South Korea accessed naver.com). Therefore, for the scope of our research, Coupang.com and 11street were chosen to retrieve customer reviews, while Naver.com was our platform of choice for obtaining blog posts. Hereafter, we briefly present the three platforms.

4.1. Coupang.com

In South Korea, Coupang.com is an e-commerce firm that, thanks to its forward-looking strategy and enormous fundings - the Japanese Softbank invested $1 billion in 2015 (Russell, 2015) - experienced a record-breaking growth. After only a few years of operation, the start-up, which was founded in 2010 by Bom Kim, a former business consultant graduated from Harvard University, has grown to become one of Amazon's major competitors in the Korean market (Dinh, 2018).

4.2. 11Street

Launched in 2008, 11Street is used by around 34

7) Alexa Confidentiality and Data Handling Overview, Amazon White Paper no. 20180720, available at https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview

8) https://myaccount.google.com/data-and-privacy

9) International Telecommunication Union (ITU) World Telecommunication/ICT Indica- tors Database

10) Source: Koreanclick.com report by Nielsen Korea

million customers. 11Street has recently made the headlines for closing an important partnership with Amazon, which will allow the platform's users to find tens of millions of Amazon US products directly on 11Street (Bae, 2020). Since its launch in 2008, every November 11th, 11Street holds its festival day which comes with discounts and offers for its clients. Just like Alibaba's Singles' Day and Amazon's Black Friday, 11Street Festival Day has become one of the major retail events in Korea. According to the data available on 11Street website, the annual event has renewed its daily merchandise volume record every year since 2017.[11)]

4.3. Naver Blog

Naver Blog is one of Naver's core services and a key blogging platform in the South Korean virtual landscape. Launched in October 2003, Naver Blog targets a wide audience, allowing its users to write and read blog posts covering any kind of topic, from business etiquette to cosmetics and electronics. Even though Naver blog only supports Korean language, the platform has rapidly become an important tool for companies willing to get in touch with their customers. In fact, companies can create a profile using Naver Blog to showcase their products and services.

The prior research examining the UK market (Maccario and Naldi, 2022a) exclusively gathered reviews from Amazon.com, as it concentrated on Amazon's Echo line, which primarily sold through that platform. In contrast, this investigation encompasses both local and global smart speakers within the South Korean market. We made efforts to compile reviews from a variety of e-commerce websites and blogs.

11) More information regarding 11Street and its Festival Day can be found at: http://global.11st.co.kr/glb/display/9885614

Ⅴ. Methodology

We performed our investigation through three stages: 1) Data collection; 2) Document selection; 3) Review Analysis. In this section, we describe the three stages.

Data collection We have collected reviews and blog posts concerning the five products described above. The reviews and posts were collected during the month of July 2021. We collected 3,448 reviews and 6,413 posts by scraping through the Python libraries BeautifulSoup and Selenium. Beautiful Soup is based on the HTML/XML analytic engine. It is a Python library commonly used for data extraction. It was developed by Leonard Richardson and other open-source developers and can parse HTML and XML documents, providing a simple way to interact with the DOM model (Richardson, 2007).

Selenium is an open-source web-based automation tool. Despite being primarily used by developers for testing tasks, Selenium has also been employed to perform data scraping and extraction and works as an add-on plugin to any browser. For our research, Google Chrome was used. We extracted the following information was extracted from each review and blog post:

- Date;
- Title of Review;
- Content.

Document selection Not all the documents collected during the first stage are of interest to our investigation. In addition, they included a lot of non-relevant words. We skimmed the documents

both to extract just the useful ones and remove the non-relevant words. As customary in any sentiment analysis task, we pre-processed the documents (reviews and blog posts) to remove all the text useless for the subsequent analysis. We also normalized the format of words to avoid any ambiguity and redundancy. The following cleaning operations were performed:

- Sentence splitting;
- Removal of numbers, white spaces, and punctuation;
- Spell check, including correction of wrong word spacing and spelling errors by using the pykospacing library in Python;
- Tokenization of sentences by using the Okt tokenizer implemented in the konlpy library in Python;
- Stopwords filtering using a pre-compiled list, which led to removing 679 of the most common Korean words, which did not have any discrimination value.[12)]

After these cleaning operations, we proceeded to extract just the documents referring to privacy issues. Following the method employed in Maccario and Naldi (2022a), the extraction stage was conducted by assessing the presence of privacy-related words. In Maccario and Naldi (2022a), the original list of privacy-related synonyms was obtained through a three steps process. First, four synonyms of the word "privacy" were identified using the Wordnet package (Miller, 1995), a lexical database used in Natural Language Processing research with applications ranging from word retrieval to machine translation (Miller and Fellbaum, 2007). Secondly, eight additional privacy-related keywords were added as a result of a semantic similarity analysis conducted through Word2Vec, an NLP technique for word representation based on neural network modeling (Mikolov et al., 2013). Lastly, fifty additional privacy-related keywords were identified by analysing the most used, inherent words, in the literature on privacy of smart speakers. In order to cope with the language employed in our Korean dataset, we had to adjust the list of keywords. The list of four privacy-synonyms obtained in Maccario and Naldi (2022a) in their first process, as described above, was translated into Korean using the online English-Korean dictionary Naver.[13)] As a result, the words *concealment*, *privateness*, *seclusion* and *secrecy* were translated as '사생활', '은폐', '개인적', '은둔' and '비밀', respectively. Then, the 51 privacy-related keywords identified in Maccario and Naldi (2022b) through a survey and the literature review were translated into Korean. In <Table 1>, we show the complete list of translated keywords employed in our research along with their English original.

After extracting the reviews of interest, we applied some validity checks. Reviews were first manually checked for duplicates. Also, some reviews were excluded because the object of the opinion was found not to pertain to the themes of security and privacy, although including one or more keywords (e.g., ' I'm worried that the one I had before broke down quickly, but so far it's working well.' contains 'worry', but it is not related to security or privacy). After performing the manual check, just twelve reviews were found to be relevant to the theme of security and privacy.

12) The stopwords list can be found at https://github.com/stopwords-iso/stopwords-ko?fbclid=IwAR39k6XIYRN7fFahioaRHu8h2-IBp-vNMShwGMR92n8xIby8LJ6O1VwjHvc

13) See the dictionary on https://en.dict.naver.com/#/main

<Table 1> English and Korean keywords

Korean Keyword	English Translation	Korean Keyword	English Translation
걱정	worry	감청	wiretapping
개인적	personal	개인	individual
공격	attack	기밀	top-secret
나타나다	peep	네트워크	network
녹음	record	누설	leak
능력	ability	도전	defiance
도청	wiretapping	몰래	stealthily
드러내다	reveal	듣기	listening
마이크	MIC	문제	problem
민감	sensitive	발견	discovery
찾음	drawing out	보이다	show
보호	protect	분석	analysis
비밀	secret	사고	accident
사생활	privacy	상호작용	interaction
송신	send	수집	collection
승인	approval	시작	start
신뢰	trust	신호	signal
악의	malice	암호	password
에이전트	agent	연결	connect
영향	effect	위험	danger
유출	outflow	유포	spread
은둔	seclusion	은폐	concealment
이해하다	understand	인식	recognize
인지	recognition	접근	access
제한	limit	조종	control
지능	intellect	지문	fingerprint
측정	measurement	침해	prejudice
클라우드	cloud	탐사	exploration
폭로	expose	프라이버시	privacy
학습	learning	iot	IOT

Ⅵ. Results

6.1. Frequency Analysis

Before performing the sentiment analysis of blog posts and product reviews to extrapolate the sentiment of Korean smart speaker users in relation to privacy, we provide a quick overview of what users talk about in their blog posts and reviews. After a manual check was performed, we only found twelve reviews that contained one or more of the selected privacy-related words. That figure is not just low on absolute terms, but it is also low in relative terms, representing a paltry 0.061% of the whole corpus (made of 9861 reviews). We can compare that result with what was obtained with the UK-based dataset analysed in (Maccario and Naldi, 2022a), which was roughly half the size of the Korean one. In that case, the number of privacy-related reviews was 2.7% of the whole corpus, which is a low figure anyway but roughly four times larger than the Korean case. <Table 2> shows the comparisons of studies for the data and methodology between the UK and South Korea.

We can immediately tick the box concerning our first research question RQ1: South Korean users buying on Coupang and 11streets, and blogging on Naver Blog are uninterested in privacy issues. The answer to RQ1 is a clear NO.

However, it is possible to obtain a wider view of the contents of the reviews and blog posts about smart speakers by analyzing the frequency of words on the whole corpus. The frequency of a word may be considered a measure of its relevance. In <Table 3>, we see the Top 10 words out of the 9861 input texts (blog and reviews), sorted by type of platform (i.e., Naver blogs vs product reviews). Similarly to what we observed in the UK, despite a general lack of interest in privacy-related issues, people appear to be concerned with sound and audio quality. The presence of the words *delivery*, *price*, and *purchase* in the top 10 suggest that users are evaluating their smart speakers based on what happened before ac-

<Table 2> Comparison of Studies between the UK and South Korea

Study	Maccario and Naldi (2022a)	This Study
Country	The UK	South Korea
Smart speakers	Amazon Echo, Echo Studio, Echo Dot	SKT NUGU Candle, SKT NUGU Mini, Kakao Mini C, Amazon Echo Dot, Google Home Mini
Number of reviews	4,756(Reviews)	9,861(Reviews and Blog posts)
Tools of scraping	Data Miner	Python – BeautifulSoup
Tools of preprocessing and sentiment analysis	R - tm,SentimentR, VADER	Python - koNLPy, KNU
Number of reviews on privacy concern	133 (2.7%)	12 (0.061%)

<Table 3> Top-10 Most Frequent Words in the Korean Corpus (by Platform)

Naver Blog			Product Reviews		
Word		Frequency	Word		Frequency
사용	Use	1764	스피커	Speaker	186
스피커	Speaker	1674	사용	Use	160
제품	Product	1153	배송	Delivery	123
제조업	Manufacturing	1053	음악	Music	122
기능	Function	945	구매	Purchase	110
연결	Connect	779	노래	Song	105
말	Speech	721	선물	Gift	105
음악	Music	684	말	Speech	100
생각	Think	605	음질	Sound Quality	96
AI	AI	604	가격	Price	95

tually using the device, e.g., the timeliness of the delivery and the high or bargain price at which they managed to buy the speaker, which echoes the findings of previous behavioural studies (Ghosh and Eastin, 2020).

6.2. Sentiment Analysis

Following the extraction of the privacy-related reviews and blog posts, we proceeded to the final two steps of our study to extrapolate the sentiment of Korean smart speaker users in relation to privacy.

- Sentiment analysis of privacy-related reviews and blog posts;
- Influence analysis of privacy opinions on the overall smart speaker evaluation.

Sentiment analysis allows us to analyse opinions, sentiments and emotions towards entities such as products, services, organizations, events, and topics, based on Natural Language Processing (NLP) techniques (Liu, 2012). However, when it comes to analyzing sentiment and opinions contained in a non-English text, a number of problems may arise.

<Table 4> Example of Positive and Negative Words from KNU Lexicon

Positive			Negative		
Korean	Translation	Score	Korean	Translation	Score
“감동”	“impressive”	2	“귀찮게 구는”	“annoying”	-1
“가치있는”	“worth it”	1	“강제로 빼앗아”	“deprive of”	-2
“감사하다”	“thank you”	2	“괴로움이나”	“suffering”	-2
“예쁘다”	“beautiful”	2	“염려가”	“concerned”	-2

As pointed out by Mohammad et al. (2016), the majority of sentiment analysis research has been conducted on English texts. There is an abundance of sentiment analysis resources for English, but a remarkable shortage of tools for many other languages.

In the context of lexicon-based approaches (and more so in their simplest version called bag-of-words approach), neither of two main approaches usually undertaken leads to reliable results (Djatmiko et al., 2019, Kaity and Balakrishnan, 2020):

- Translating the focus-language text into a resource-rich language such as English to employ a powerful English sentiment analysis system on the translated text;
- Translating sentiment analysis resources such as labelled corpora and lexicons from English into the focus language and applying them to perform sentiment analysis in the focus language.

In order to overcome the risks connected to these two approaches, we avoided translation altogether for sentiment analysis purposes, and we directly employed the KNU Korean sentiment lexicon, performing the whole analysis inside a Korean framework (Park et al., 2020). KNU Korean Sentiment Lexicon is a general-purpose Korean dictionary, which consists of sentiment tokens obtained by analyzing the entries contained in the Standard Korean Language Dictionary. KNU is based on a lexicon of 14854 tokens, which can be unigrams, bigrams, phrases, or sentence patterns. Each token is assigned an individual score in the range [-2;2], where the zero value represents a neutral sentiment. The closer the value is to 2, the more positive the sentiment is. In <Table 4>, some of the most common positive and negative Korean words appearing in KNU Lexicon are presented along with their sentiment score and English translation.[14] Although having been launched in 2018, the KNU package has already been employed in several papers. A recent application to Korean social media data can be found at Shim et al. (2021).

The results obtained with KNU are reported in <Table 5>, showing that seven reviews obtained an overall positive sentiment score, with review no.8 marking the highest sentiment score (11). Instead, reviews no. 1, 4 and 11, were found to carry a negative sentiment, showing a sentiment score of -6, -4, and -1, respectively. Sentence sentiment show sentiment scores of the sentences containing the privacy keyword of <Table 1>. Though the sentences contain the keywords, their sentiments seem neutral. Review sentiment average is the average sentiment score of sentences in the reviews.

The followings are three example reviews related to privacy concerns:

14) The KNU Korean Sentiment Lexicon is available at: https://github.com/park1200656/KnuSentiLex

- No. 1: There's a feature that reads KakaoTalk messages out loud, but since we're going to use it together in the living room, we didn't use that privacy feature.
- No. 3: Echo Dot may not be a great purchase for thriving smart home enthusiasts, but the Kids Edition is highly recommended for parents who want a little more privacy for their young children.
- No. 4: By touching the center of the Google Home Mini, you can pause music or audio content, but recently it was revealed that this feature was automatically recording and transmitting the surrounding environment to Google servers, regardless of the user's intention. This caused privacy concerns and as a result, Google permanently disabled this feature.

The results obtained with KNU revealed varying sentiments towards privacy. Review no. 1 exhibited a neutral sentiment, as the user acknowledged the privacy feature but chose not to use it. Review no. 3 displayed a positive sentiment, as the user recommended the Kids Edition for parents who want more privacy for their children. Review no. 4 had a negative sentiment, as the user expressed concern about the privacy breach caused by the Google Home Mini feature.

In the previous section, we provided an answer to RQ1. We now address the research questions RQ2 and RQ3. The answer to RQ2 agrees with the findings from past research in that privacy appears to be a controversial subject: three reviews were found to carry a negative sentiment, seven reviews were reported to be positive, and lastly, reviews no. 2 and no. 6 were neutral. As for RQ3, we have computed the correlation between the sentiment scores obtained in privacy-related sentences vs the whole reviews

<Table 5> Sentiment Score in Privacy-related Reviews and Privacy-related Sentences (KNU package)

Element ID	Sentence Sentiment	Avg Review Sentiment	Review Sentiment
1	0	-0.6	-6
2	0	0	0
3	2	0.2	3
4	-2	-0.333	-4
5	1	0.5	2
6	0	0	0
7	0	0.25	3
8	0	1.1	11
9	-1	0.4375	7
10	0	0.2222	2
11	0	-1	-1
12	0	0.125	1

to which those sentences belonged. For that purpose, we employed the Pearson correlation coefficient between the sentiment scores obtained for the two groups.

We obtained a correlation coefficient of 0.6227, which cannot be considered very high, though positive. As the blog posts have no rating scores, we did not compute the correlation coefficient between ratings and sentiment scores. We can then answer positively to RQ3. Though very few customers are concerned about privacy, their sentiment towards privacy positively influences the overall sentiment about smart speakers in South Korea, though the influence is not extremely high.

Ⅶ. Discussin and Conclusions

With the goal of substantiating the findings from Maccario and Naldi (2022a) and widening the search scope to include different devices and, most im-

portantly, different countries where the cultural attitude towards privacy may be different, we chose to replicate and adapt a text-mining analysis on smart speakers in South Korea, i.e., a well-developed Asian country vs a Western country as examined in Maccario and Naldi (2022a). With its impressive rate of diffusion of innovation and competitive internal smart speakers' market, South Korea can be considered an ideal target country for our study. Internationally, Amazon and Google hold prominent positions in the smart speaker market, with numerous Chinese brands seeking to increase their market shares. As depicted in <Figure 2>, the South Korean market exhibits a distinct landscape compared to global competition. Key players in this region include Korean Mobile Network Operators such as SKT and KT, as well as IT Tech companies like Kakao and Naver, which offer search engine and messenger services. Local brands account for approximately 60% of the market share. Notably, the privacy features exhibited by local and global brands are quite similar. These brands allow users to disable the microphone and necessitate wake words to initiate voice recording and conversation. Furthermore, they offer privacy settings for users to opt out of sharing their voice recordings and to request deletion of the recorded data.

With the aim of investigating the relevance of privacy concerns in smart speaker product reviews and blog posts in South Korea, three research questions were posed. It was subsequently found that South Korean smart speaker users seem to be either unconcerned or unaware of privacy issues. Furthermore, even those few reviews that mention the topic were not found to carry a negative sentiment unequivocally. However, the sentiment towards privacy by those few users positively influences their overall sentiment towards smart speakers.

Compared to previous studies on the Western market (Maccario and Naldi, 2022a), this research corroborates the limited interest in privacy issues while demonstrating an even lower level of interest in an Asian country as opposed to a Western one (see <Table 2>). The UK study categorized privacy concerns into four groups: Audio features, Security Issues, Data collection and transfer, and Device intelligence. This study revealed that the primary privacy concerns in reviews relate to Data collection and transfer, with users expressing apprehension about voice recordings and potential unauthorized leaks. Despite the presence of privacy concerns in customer reviews from both the UK study and this investigation, the sentiments were not overly negative; in fact, most sentiment scores were positive. While the areas of concern and the degree of worry differ slightly, the impact of privacy concerns on review sentiment remains notably similar between the two countries. These findings underscore the necessity of considering cultural differences when analyzing product reception across different nations.

This study holds academic implications for broadening research on the quantitative evidence of users' privacy considerations. Firstly, this paper adopts the research methodology of the previous study (Maccario and Naldi, 2022a) while factoring in local competition among smart speaker companies. By gathering customer reviews from various brands based on South Korea's market share and from multiple e-commerce sites and a blog company, this approach offers a more comprehensive perspective on user concerns compared to focusing on a single brand and e-commerce site. Secondly, this study extends the previous research to a different country with a distinct language. We translated privacy keywords as shown in <Table 1> and employed a lexicon-based sentiment analysis for the Korean language.

Sentiment analysis tools, such as VADER and SentimentR, often do not cover Asian languages. This expansion could serve as a model for other studies aiming to extend their geographical scope. The methodology presented in this study offers managerial implications for smart speaker sellers and manufacturers. In response to concerns about data breaches, most brands now incorporate microphone-off and privacy settings to mitigate unwanted data transfer. However, privacy features are not always prominently displayed on product pages or even within service and product manuals. Although many customers may not express significant concern about smart speaker privacy issues in their reviews, this does not imply they are unconcerned about privacy overall. Privacy concerns remain crucial for most digital devices. As reviews mentioning privacy issues often highlight privacy features, it is advisable for companies to clearly explain these features.

In this paper, we acknowledge several limitations in our investigation. First, we have focused on a single country, which cannot adequately represent the entire Asian continent, just as the UK cannot be considered representative of the whole Western world. Second, our analysis is limited to smart speaker owners, who have already expressed their positive attitude towards the products by purchasing them. As a result, the sample can be considered biased in favor of the product. The act of buying a smart speaker also indicates a level of tolerance towards its potential drawbacks, including privacy concerns. Additionally, we recognize the constraint posed by the limited number of reviews addressing privacy concerns in this paper. While our collection of reviews surpassed that of prior work, we obtained only a small number of reviews specifically concerning privacy. We contend that acquiring sufficient data for a robust statistical analysis remains challenging even with an increased number of crawled reviews. Consequently, the paper's analysis primarily involves a rudimentary sentiment assessment and correlation calculations.

Acknowledgements

Hong Joo Lee was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A3A2A02093277), and by The Catholic University of Korea, Research Fund, 2023.

These authors contributed equally to this work.

<References>

[1] Abdi, N., Ramokapane, K. M., and 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 (SOUPS 2019)* (pp. 451-466).

[2] Ackerman, M. S., Cranor, L. F., and Reagle, J. (1999). Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM Conference on Electronic Commerce* (pp. 1-8).

[3] Bae, I. Y. (2020). Amazon partners with Korean e-commerce platform 11Street. Retail in asia. Retrieved from https://retailinasia.com/in-trends/amazon-partners-with-korean-e-commerce-platform-11street

[4] Bawack, R. E., Wamba, S. F., and Carillo, K. D. A. (2021). Exploring the role of personality, trust, and privacy in customer experience performance during voice shopping: Evidence from sem and fuzzy

set qualitative comparative analysis. *International Journal of Information Management, 58,* 102309. https://doi.org/10.1016/j.ijinfomgt.2021.102309

[5] Bellman, S., Johnson, E. J., Kobrin, S. J., and Lohse, G. L. (2004). International differences in information privacy concerns: A global survey of consumers. *The Information Society, 20*(5), 313-324. https://doi.org/10.1080/01972240490507956

[6] Bolton, T., Dargahi, T., Belguith, S., Al-Rakhami, M. S., and Sodhro, A. H. (2021). On the security and privacy challenges of virtual assistants. *Sensors*, *21*(7), 2312. https://doi.org/10.3390/s21072312

[7] Cho, H., Rivera-S´anchez, M., and Lim, S. S. (2009). A multinational study on online privacy: global concerns and local responses. *New Media & Society*, *11*(3), 395-416. https://doi.org/10.1177/1461444808101618

[8] Choi, J., Lee, H. J., Sajjad, F., and Lee, H. (2014). The influence of national culture on the attitude towards mobile recommender systems. *Technological Forecasting and Social Change*, *86*, 65-79. https://doi.org/10.1016/j.techfore.2013.08.012

[9] Clark, J. W., Snyder, P., McCoy, D., and Kanich, C. (2015). "i saw images i didn't even know i had" understanding user perceptions of cloud storage privacy. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1641-1644).

[10] Dinh, K. P. (2018). The internationalization strategies of e-commerce SMEs and start-ups in South Korea: Case studies of four e-commerce start-ups: Cconma, coupang, memebox, ticketmonster. *Bachelor's Thesis*, Aalto University.

[11] Djatmiko, F., Ferdiana, R., and Faris, M. (2019). A review of sentiment analysis for non-english language. In *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)* (pp. 448-451). IEEE.

[12] Dubois, D. J., Kolcun, R., Mandalari, A. M., Paracha, M. T., Choffnes, D., and Haddadi, H. (2020). When speakers are all ears: Characterizing misactivations of iot smart speakers. In *Proceedings on Privacy Enhancing Technologies*, 2020(4).

[13] Ghosh, C., and Eastin, M. S. (2020). Understanding users' relationship with voice assistants and how it affects privacy concerns and information disclosure behavior. In *International Conference on Human-Computer Interaction* (pp. 381-392). Springer.

[14] Hofstede, G. (2003). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations* (2nd ed.). London: SAGE Publications.

[15] Jiang, X., and Apthorpe, N. (2021). Automating internet of things network traffic collection with robotic arm interactions. *arXiv preprint arXiv: 2110.00060*. https://doi.org/10.48550/arXiv.2110.00060

[16] Kaity, M., and Balakrishnan, V. (2020). Sentiment lexicons and non-english languages: A survey. *Knowledge and Information Systems, 62*(12), 4445-4480. https://doi.org/10.1007/s10115-020-01497-6

[17] Knijnenburg, B. P., Page, X., Wisniewski, P., Lipford, H. R., Proferes, N., and Romano, J. (2022). *Modern Socio-Technical Perspectives on Privacy.* Springer International Publishing.

[18] Lin, L., Liu, X., Fu, X., Luo, B., Du, X., and Guizani, M. (2021). A non-intrusive method for smart speaker forensics. In *ICC 2021-IEEE International Conference on Communications* (pp. 1-6). IEEE.

[19] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1-167.

[20] Maccario, G., and Naldi, M. (2022a). Alexa, is my data safe? the (ir) relevance of privacy in smart speakers reviews. *International Journal of Human–Computer Interaction*, *39*(6), 1244-1256. https://doi.org/10.1080/10447318.2022.2058780

[21] Maccario, G., and Naldi, M. (2022b). Privacy in smart speakers: A systematic literature review. *Security and Privacy*, *6*(1), e274. https://doi.org/10.1002/spy2.274

[22] Malin, C. (2019). Home-grown innovation driving South Korea's voice assistant market, Asia AI News.

Retrieved from https://medium.com/@asiaainews/home-grown-innovation-driving-south-koreas-voice-assistant-market-aea64c36340d

[23] Mandalari, A. M., Kolcun, R., Haddadi, H., Dubois, D. J., and Choffnes, D. (2020). Towards automatic identification and blocking of non-critical iot traffic destinations. *arXiv preprint arXiv:2003.07133*. https://doi.org/10.48550/arXiv.2003.07133

[24] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems, 26*. https://doi.org/10.48550/arXiv.1310.4546

[25] Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM, 38*(11), 39-41. https://doi.org/10.1145/219717.219748

[26] Miller, G. A., and Fellbaum, C. (2007). Wordnet then and now. *Language Resources and Evaluation, 41*(2), 209-214. https://doi.org/10.1007/s10579-007-9044-6

[27] Mohammad, S. M., Salameh, M., and Kiritchenko, S. (2016). How translation alters sentiment. *Journal of Artificial Intelligence Research, 55*, 95-130. https://doi.org/10.1613/jair.4787

[28] Neville, S. (2020). Eavesmining: A critical audit of the amazon echo and alexa conditions of use. *Surveillance & Society, 18*(3), 343-356. https://doi.org/10.24908/ss.v18i3.13426

[29] Nisbett, R. E. (2003). *The Geography of Thought: How Asians and Westerners Think Differently...and Why*. New York: The Free Press.

[30] Park, H. M., Kim, C. H., and Kim, J. H. (2020). Generating a korean sentiment lexicon through sentiment score propagation. *KIPS Transactions on Software and Data Engineering, 9*(2), 53-60. https://doi.org/10.3745/KTSDE.2020.9.2.53

[31] Pridmore, J., and Mols, A. (2020). Personal choices and situated data: Privacy negotiations and the acceptance of household intelligent personal assistants. *Big Data & Society, 7*(1). https://doi.org/10.1177/2053951719891748

[32] Richardson, L. (2007). Beautiful soup documentation. Retrieved from https://www.crummy.com/software/BeautifulSoup/ bs4/doc/

[33] Russell, J. (2015). A softbank invests $1B in Korean E-commerce firm Coupang at a $5B Valuation. *Techcrunch*. Retrieved from https://techcrunch.com/2015/06/02/coupangzillion/

[34] Shim, J. G., Ryu, K. H., Lee, S. H., Cho, E. A., Lee, Y. J., and Ahn, J. H. (2021). Text mining approaches to analyze public sentiment changes regarding covid-19 vaccines on social media in Korea. *International Journal of Environmental Research and Public Health, 18*(12), 6549. https://doi.org/10.3390/ijerph18126549

[35] Sudharsan, B., Corcoran, P., and Ali, M. I. (2019). Smart speaker design and implementation with biometric authentication and advanced voice interaction capability. In *AICS* (pp. 305-316).

[36] Tabassum, M., Kosinski, T., and Lipford, H. R. (2019). "i don't own the data": End user perceptions of smart home device data practices and risks. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)* (pp. 435-450).

[37] Yeo, J. (2019). Smart speaker privacy concerns spread to Korea. The Korea Herald. Retrieved from https://www.koreaherald.com/view.php?ud=20190909000880

[38] Zhang, N., Mi, X., Feng, X., Wang, X., Tian, Y., and Qian, F. (2019). Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems. In *2019 IEEE Symposium on Security and Privacy (SP)* (pp. 1381-1396).

◆ About the Authors ◆

Hong Joo Lee

Hong Joo Lee is a Professor of Business Administration at The Catholic University of Korea. He has a Ph.D. from the KAIST Business School (2006) and was with the MIT Center for Coordination Science as a postdoctoral fellow in 2006. He stayed at the MIT Center for Collective Intelligence as a visiting scholar in 2011 and was an international researcher at the Brunel Business School, Brunel University in 2017－2018. His research areas are utilizing intelligent techniques and analyzing the effects of intelligent aids in digital business. His papers have been published in International Journal of Electronic Commerce, Decision Support Systems, Expert Systems with Applications, Journal of Electronic Commerce Research, Cyberpsychology, Behavior and Social Networking, Information Systems Frontiers, and other journals.

Guglielmo Maccario

Guglielmo Maccario is a Contract Professor of Economics of Globalization and Development Economics and International Cooperation at UNINT University in Rome. He has a Ph.D. from UNINT University (2022). He is currently working as Data Analyst at the European Central Bank ECB where he carries out research and analysis for the Language Analytics team. Guglielmo Maccario's research interests encompass innovation, entrepreneurship, digital divide, and the application of natural language processing (NLP) techniques to various aspects of these areas.

Maurizio Naldi

Maurizio Naldi is a Full Professor of Computer Science at LUMSA University in Rome. He received his MSc degree in electronic engineering from the University of Palermo, Italy, in 1988 magna cum laude and his PhD degree in telecommunications engineering from the University of Rome Tor Vergata, Italy, in 1998. From 1988 to 2000, he pursued an industrial career, working in the R&D departments of several ICT companies, ending as Head, Traffic Forecasting and Network Cost Evaluation Group. In the same period, he acted as Rapporteur in ITU and ETSI standardization bodies. He moved to academia in 2000, first as an Assistant Professor and then as an Associate Professor at the University of Rome Tor Vergata till 2019. He's Co-Editor of Electronic Commerce Research and Applications (Elsevier) and is on the Editorial Board of the Journal of Business Research, PLOS One, and the Journal of Universal Computer Science (J.UCS). His research interests span the fields of network and service economics as well as the applications of statistical decision theory and machine learning.

Submitted: January 26, 2023; 1st Revision: April 4, 2023; Accepted: August 30, 2023