

# ETHICS OF SINGING VOICE SYNTHESIS: PERCEPTIONS OF USERS AND DEVELOPERS

Kyungyun Lee<sup>1</sup>

Gladys Hitt<sup>2</sup>

Emily Terada<sup>2</sup>

Jin Ha Lee<sup>2</sup>

<sup>1</sup> Gaudio Lab, Seoul, Korea

<sup>2</sup> University of Washington, Seattle, USA

mo@gaudiolab.com, hittg@uw.edu, eterada@uw.edu, jinhalee@uw.edu

## ABSTRACT

Singing Voice Synthesis (SVS) has recently garnered much attention as its quality has improved vastly with the use of artificial intelligence (AI), creating many opportunities for supporting music creators and listeners. Recently, there have been growing concerns about ethical issues related to AI development in general, and to AI-based SVS development specifically. Many questions remain unexplored about how to ethically develop and use such technology. In this paper, we investigate the perception of ethical issues related to SVS from the perspectives of two different groups: the general public and developers. We collected 3,075 user comments from YouTube videos showcasing various uses of SVS as part of a mainstream variety show. Additionally, we interviewed six researchers developing SVS technology. Through thematic analysis, we identify and discuss three different aspects related to ethical issues in SVS development, highlighting the similarities and differences between the perspectives of the general public and developers: (1) Use scenarios, (2) Attitudes towards development, and (3) Meaning of "Creativity", and (4) Concerns about human rights, intellectual property (IP) and legal issues.

## 1. INTRODUCTION AND BACKGROUND

Artificial intelligence (AI) is developing rapidly and becoming increasingly pervasive in our lives through various forms of technology, including robots and smart devices. The AI field has been thriving over the past decade due to the increasing availability of massive data, computational resources, and deep-learning-based architectures [1, 2]. Related to music, there are various AI technologies to support music organization, production, and enjoyment, including recommendation [3, 4], music generation [5, 6], instrument sound synthesis [7] and singing voice synthesis [8–10]. Singing Voice Synthesis (SVS) has existed since the late 1950s [11] and has been used and enjoyed by many users through applications like singing synthe-

sizers and vocaloids [8]. The significant improvement in the quality of voices due to the use of recent techniques, specifically deep learning, has given rise to more opportunities for the use of SVS, but has also led to more questions regarding its ethical use and increased concern for potential misuse. While the SVS community has started discussing some of the ethical issues related to the development and use of SVS, the discussion is in the early, nascent stage [2]. Furthermore, there is limited research investigating how the general public, as potential users, perceive and feel about the use of SVS, especially since the synthesized singing voices are sounding increasingly realistic and almost indistinguishable from human voices.

The ethical issues encompass multiple areas, including intellectual property (IP) (e.g., the ownership of the media with AI voices) and human rights (e.g., who decides how the voices are used). In addition to understanding how the general public thinks about SVS, it is important to consider what influences their perceptions and whether developers are aware of them. If developers are aware, does that impact their goals and direction? Understanding how the general public perceives AI technology may offer useful insights to developers on how to address negative perception and ensure AI is well-received by users. It is not only important for envisioning the potential commercialization of such technologies, but also for discussing ethical issues and collaboratively thinking about how society should address such issues.

In this paper, we aim to improve our understanding of how people perceive the uses of SVS technology through a content analysis of user comments collected from on-line videos showcasing various applications of SVS. To examine how user perception on ethical issues related to SVS can influence the development of the technology and vice versa, we also interviewed SVS developers regarding viewpoints on potential ethical issues. We aim to answer the following research questions: (1) How does the general public react to AI-generated singing voices, and what are the implications to SVS developers? (2) What kinds of ethical issues do users and SVS developers consider?

## 2. LITERATURE REVIEW

### 2.1 Music-related AI Technologies and SVS

With the advancement of generative models in deep learning, creative applications of music AI technology have be-



© K. Lee, G. Hitt, E. Terada, and J. H. Lee. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).  
**Attribution:** K. Lee, G. Hitt, E. Terada, and J. H. Lee, "Ethics of Singing Voice Synthesis: Perceptions of Users and Developers", in *Proc. of the 23rd Int. Society for Music Information Retrieval Conf.*, Bengaluru, India, 2022.

come a popular research topic in music information retrieval (MIR). Music generation is one area, encompassing various tasks, such as melody generation [5], expressive performance generation [12], sound synthesis [7] and pop song generation [6]. Research in generative music not only varies in tasks, but also in their data formats, which can be at a symbolic-level (e.g., scores and MIDI tracks), at an audio-level (e.g., mp3 and wav files) or a mix of both. In the past years, music AI technologies have become increasingly available and accessible to the public. One example is the BachDoodle, a Bach-like harmonization system from Google Magenta [13], which received much attention when made publicly available as an interactive Google Doodle<sup>1</sup>. Another example is an audio-level pop music generation system, called Jukebox, which generates raw audio of pop songs, including singing voices [6]. This system is conditioned on lyrics and artists, therefore allowing control over such features to obtain desired outcomes. The quality of generated music has vastly improved, even motivating musicians, such as The Flaming Lips, to experiment with AI systems in their music creation process<sup>2</sup>.

SVS has been researched for more than two decades. The goal of SVS is to learn human-like singing voices given input conditions, like speech, lyrics and melodies. A thread of research focuses on creating non-existent voices as seen in popular applications like vocaloids [14]. Other areas handle existing voices: replicating existing singing voices [9,10] such as those of popular artists, correcting the singing expressions for performance improvement [15], and converting speech to singing voice [16]. We have observed significant improvements in the synthesis quality through the adaptation of advanced deep learning models, such as Tacotron [17] and DeepVoice [18]. In particular, the YouTube videos analyzed in this paper replicated voices of popular singers from music scores and lyrics [9], incorporating techniques from text-to-speech (TTS) with adversarial training to achieve state-of-the-art results.

## 2.2 Ethical Issues Related to Music AI

Concerns regarding intellectual property and AI are growing. Many are concerned about how to protect artists' value in the music industry when "automation" may drive music production costs to zero [19]. While we may still doubt the quality of AI musicality, SKYGGE's Hello World<sup>3</sup> and virtual musicians such as Vocaloid Hatsune Miku<sup>4</sup> are clear signs that the "trend to autonomous and human-like virtual musicians is extremely well established" [20]. As such, ethical concerns are also tied to concerns over excessive loss in human creativity [14]. There is also an aspect of insufficient public knowledge about these technologies, which in turn, informs the lack of legal regulations. Furthermore, such regulations would be heavily influenced by corporations which have "incentives to ignore public input and shift regulation to their benefit" [21]. According

to Floridi, there are five digital ethics pitfalls that can be applied when examining case studies and potential consequences of SVS/AI technologies: 1) Ethics Shopping, 2) Ethics Bluewashing, 3) Ethics Lobbying, 4) Ethics dumping, 5) and Ethics Shirking [22].

Beyond the ethics of artist getting compensated for their voices or work, there are also additional questions with regards to human rights – should artists be allowed to "refuse" having their voices or work serve as data sets for generating new voices, and potentially a third party profiting from it [23,24]?

Existing literature emphasizes the importance of drawing ethical considerations and conclusion within the context of AI's socio-technical systems, rather than on the specific features of AI itself as today's digital technology "ecosystem" necessitates the inclusion of more stakeholders and perspectives [25]. Research finds that Western perspectives often dominate the conversation surrounding AI [26–28] and increasingly emphasizes the importance of considering cultural differences when examining human-AI interactions and attitudes towards AI [29,30]. There is a pressing need for specific applications of ethics and accountability measures rather than current abstract principles [26] with the argument that stakeholders in the music ecosystem must be the ones responsible for establishing an ethical response for the purpose of a sustainable music industry that includes humans and non-humans [31].

## 3. STUDY DESIGN AND METHODS

To understand users' and developers' perspectives on AI-based SVS technology, we employed a mixed method approach using thematic analysis of online user comments and interviews. Given that there is currently no widely distributed commercial application using this technology (other than vocaloid communities which are generally considered to be niche), it was challenging to identify a specific community to represent the perception of the general public. Inviting people to react in an experimental setting also has the potential for increased participant response bias [32]. Therefore we opted to find data sources which would allow us to capture user reactions in a more natural setting. We looked into public online videos showcasing state of the art technology and found promising examples from Korean TV shows called "AI versus Humans"<sup>5</sup> and "One More Time,"<sup>6</sup> where several different use cases for this technology were presented, such as generating voices of artists (both currently active and deceased), manipulating the artists' voice, generating singing voices in languages different from the language of the original voice samples, and joint performance of artist and AI. A total of 3,075 comments were collected from eight videos from 2020 to 2021 where they presented various applications of AI-based SVS technology.

Most comments were in Korean and a few in English. Two authors who are fluent in Korean independently coded

<sup>1</sup> <https://www.google.com/doodles/celebrating-johann-sebastian-bach>

<sup>2</sup> <https://magenta.tensorflow.org/fruitgenie>

<sup>3</sup> <https://www.skygge.fr/>

<sup>4</sup> <https://ec.crypton.co.jp/pages/prod/virtualsinger/cv01>

<sup>5</sup> <https://programs.sbs.co.kr/enter/aivshuman/about/67201>

<sup>6</sup> <https://program.genie.co.kr/onemoretime/main>

these comments. Taking an inductive approach, the two authors reviewed partial data, used Mural board (a collaborative online tool) to organize the concepts, and came up with the initial codebook through a thematic analysis [33]. Afterwards, the authors coded sample comments together and iterated on the design of the codebook based on discussion. These discussions resulted in addition of a new code (Positive tech-Other) and redefinition of "Commercial" and "Personal" for a clearer distinction. The two coders then used the final codebook (<https://osf.io/7em95/>) to code all the comments following a consensus model [34], and discrepancy in code application was discussed with the goal of reaching a consensus.

For additional data on developers' perspectives, we reached out to authors of recent publications on this technology asking whether they would participate in an online interview. Each interview was recorded via Zoom and verbal consent was obtained at the start of recording, following the protocol approved by the UW Institutional Review Board. Using Zoom to host and record interviews also enabled us to generate interview transcripts which the authors cleaned and coded for analysis. Interview questions focused on the kinds of ethical issues developers consider, the broader implications to designers of music AI technology, how they perceive the general public's reaction to music AI technology, and what ethical and legal precautions they believe should be implemented.

We reached out to 12 authors, and were able to recruit six developers from three different countries. Five developers were faculty or students building and testing the AI-based SVS technology, and one was an artist who provided their voice samples to build the data set for this technology. Since there is a limited number of researchers who work on this particular technology, the overall pool of users is smaller compared to the number of general AI researchers. The small sample size makes this exploratory in nature.

All interviews were fully transcribed and coded using an inductive approach [35]. Two authors created the initial codebook through thematic analysis, following a similar process as above. The final codebook had 16 codes. Using a qualitative coding software ATLAS.ti, the four authors coded the interviews. We assigned two different coders to each interview, and followed the consensus model [34].

## 4. FINDINGS

### 4.1 Perception of the General Public

Eight categories and 27 associated codes emerged from our analysis of online user comments from YouTube videos. "Positive emotions" (e.g., awe) and "Negative emotions" (e.g., fear) categories include responses that reflect users' emotions evoked by seeing the application of SVS. We also created "Positive tech" and "Negative tech" categories that consist of comments about how good or limited they think the current SVS technology is. We dedicated a category, "Conflicted", for comments that showed users' mixed feelings on whether their current emotion is justifiable or not. "Considerations" category contains codes representing different aspects related to ethical issues (e.g., copyright, hu-

man right). "Challenges" contains codes that mention realistic problems related to SVS and AI (e.g., misuse). Lastly, "Opportunities" includes comments that suggest new possible use cases and applications of the current SVS technology (e.g., commercial, personal). Reviewing 3,075 comments, excluding incomprehensible ones, resulted in 1,190 comments that were coded. The result of our analysis as a code distribution can be accessed at: <https://osf.io/7em95/>.

Overall, we observed positive and negative sentiments in a 65% to 35% ratio. The most observed comment, coded as "awe" (130/1,190), conveyed surprise and fascination towards the level of today's SVS technology. Along with "awe", "moving" (104) and "nostalgia" (84) were commonly appearing positive themes. Several users were touched by the revival of voices of the singers they grew up with (e.g., *"I am thankful for the opportunity to listen to [the artist's] voice again"*). The particular use of the technology did seem to influence audiences' emotion (e.g., nostalgia for deceased artists). The high frequency of the code "moving" implies the potential power and influence of AI on humans. Negative and cynical comments about SVS and AI were represented by codes "fear" (81), "guilt" (25) and "neg-emotions-other" (19). We encountered comments, such as *"this is giving me goosebumps"*, expressing shock and even discomfort towards the quality of reenactments of familiar voices. Feelings of "guilt" were also represented - *"isn't this humiliating the deceased?"*, *"I don't think it's appropriate to do this"* - showing irritation and criticizing lack of morality and respect towards the deceased artists. We observed some conflicted feelings - *"I don't know if I feel good or scared about the regeneration of artists' voice. I feel resistance, as well as curiosity at the same time. I don't know"*, *"It feels weird...I feel scared. Maybe I shouldn't have listened to this"* and *"I do want to see the artist, but not in this way."* Regarding the aspect of technical advancements, there were comments both praising and belittling the results. Some were impressed and rated AI as sounding exactly like the original artists-*"(AI) even replicated the way (the artist uniquely) pronounces 'r'"* - with a few users thinking the AI even sounded better than the original artists.

The "Opportunities" category included comments about audiences' desire to use the presented SVS technology for their personal interest. There were series of comments asking for the regeneration of the voices of other artists or their family members - *"Can they also replicate X?"* and *"I wish to hear my father's voice."* Several comments requested the commercialization of the SVS technology, expressing interest in purchasing the application if made available. There were also comments discussing other potential use cases, such as voice actors in films or saving their own voices to sing songs to their grandchildren posthumously.

From the prominence of the code "misuse," we witnessed growing concerns towards AI in today's society. Often referencing deepfakes, users were afraid of the rise of potential scams (e.g., voice phishing). Some even expressed anger at the irresponsible development of such "dangerous" technology, arguing that it will do more harm

than good. Along this line of negativity, dystopic comments expressed concerns about the potential of AI to replace humans-*"I am a vocal student...Help me, I think my job is disappearing"*, *"I wonder how many jobs will be lost to AI in the near future."* These comments led to sentiments of inevitability and reluctant acceptance- *"we need to make ourselves irreplaceable by AI."* Meanwhile, we observed only one comment discussing the ways to overcome the potential dangers of the technology ("counter").

As reflected in the code "guilt", some addressed issues regarding the legality of SVS usage in terms of human rights and copyright. Many considered it disrespectful and unethical to use voices without the actual owner's consent, invalidating families' and relatives' compliance. They were also critical of the notion that the show could be profiting from the exhibition of deceased artists' voices.

## 4.2 Perception of Developers

Six interviewees were highly involved in the development of SVS technology, with experience ranging from two to 20 years. Their initial motivation driving their research was curiosity about the technology. All participants envisioned their work potentially helping musicians increase their productivity and creativity. For example, they envisioned SVS being used to quickly generate demo tracks for vocalists, as it is inefficient to invite the vocalists each time to record demos (P1, P5). While participants unanimously predicted that the current SVS technology will reach the level of human voices within the next five to 10 years, they also listed limitations, including data constraint and the need for higher quality results for commercial applications. Developers' thoughts on the "meaning of creativity" were related to the limited data set. They questioned the possibility of AI becoming truly creative, as *"AI can only generate average of the data"* and *"AI is not able to generate something better than humans since it is only trained on the input data"* (P1, P4). P1 mentioned that artists incorporate context and history into the music they are making, which AI lacks. P6 pointed out that the well-known vocaloid Hatsune Miku's fandom is not just towards the virtual character itself, but also towards the human creators behind it.

On the discussion about ethical issues of SVS, a prevalent response was that developers are aware of such issues and their importance, but no one knows specific ways to handle them. Debates have arisen around the unclear standards and practices regarding ownership problems in the music industry. Regarding who is responsible for addressing ethical issues, some participants emphasized that developers should take more caution when it comes to data use and how they make their technology available to the public (P4), while others suggested that there should be a group of experts dealing specifically with ethical issues so the developers can focus solely on the advancement of the technology (P3). However, half of the developers noted that the SVS technology of today is not yet advanced enough to consider ethical issues seriously (P2, P3, P4).

Regarding how to address the ethical and legal impli-

cations of this technology, participants indicated that solutions to counteract the misuse of AI are necessary. For example, most participants (all but P5) mentioned the development of counter technologies which can detect AI generated voices as one of the solutions to the potential misuse, and were fairly confident in the power of counter technologies. In general, participants maintained positive views on the future of SVS and considered the general public's negative reactions towards AI to be no different than similar reactions towards other technologies in the past. P5, for instance, talked about the initial negative reactions towards the sound of electronic keyboards and how those attitudes changed as keyboards become more widely used in the music industry. They noted that, similarly, the advancement of AI is inevitable and the general public will slowly accept it. Upon closing the interviews, participants commented on the future of SVS technology, ranging from concerns - *"I'm not sure if it's okay to get comfortable with the mass production of voice through AI"* (P1)-to practical directions to take- *"Humans should try to figure out ways to co-exist with AI, not compete with or hinder its development, because AI will never be able to replace humans"* (P2)-to the need for developers, artists, and the public to maintain an open line of communication regarding differing motivations (P4,P6).

## 5. DISCUSSION AND IMPLICATIONS

### 5.1 Perspectives on Use Scenarios

All of the developers interviewed considered supporting creators including composers, performers, and producers, as one of the main goals of their research. When asked to share what kinds of use scenarios they envisioned for the application of their work, they discussed various situations in which AI supports and benefits the creators: *"a tool for the creator like autotune or mixing"* (P2), a tool for generating demo tracks for vocalists (P1), and a means of *"style transfer to correct and improve singing"* (P4). P6, in particular, emphasized that mimicking the human voice to make it sound "natural" is only one aspect of SVS and the true potential lies in generating a variety of voices. They explained how some users may want the voice to sound "artificial" and even prefer that, as exemplified by the popularity of Vocaloid's Hatsune Miku, and the frequent use of Autotune in mainstream music.

The developers also tended to lean towards a more controlled model where a select few have direct access to the use of technology or data set to prevent potential misuse. A question was raised about the commercialization of SVS and its implications. P2, in particular, explained using SVS to recreate the voices of existing or deceased artists will have limitations due to ethical issues beyond just using it as a proof-of-concept or for an event, but generating new voices might have more freedom to be used commercially.

However, the user comments demonstrate a range of desires and ideas for different kinds of commercial uses for SVS technology. Of the 62 comments mentioning potential commercial uses, many expressed that they would be interested in purchasing a song or album by a deceased artist using SVS. Some even envisioned specific apps peo-

ple could download and use like *"a paid app that lets us pick the singer's voice and the song we want to hear in that voice. Maybe in our own voice!"* or *"a business with streaming websites that lets people pay and save the songs.* Many users also imagined the use of the SVS technology beyond the context of music, such as using *"this technology to speak with dead people"*. A user wondered about their future with this technology in relation to human interactions across time (*"Even if I die early, I can sing a lullaby to my grandchildren?"*).

Users also had more diverse ideas about the potential misuse of the technology, including ethical and legal issues that could stem from its development. While this could potentially be attributed to the larger number of comments seen on YouTube compared to those gathered from the interviews, it does indicate that users are definitely considering the commercialization of this technology and are willing to pay for and use it for "personal" goals.

## 5.2 Attitudes Towards AI Technology Development

Developers tended to be interested in and focused on enhancing the current technology, often recognizing the potential misuse or abuse of technology while accepting the reality that the technology will continue to improve over time regardless of how they feel or act. Despite their initial interest in the research being motivated from the excitement about the technology itself (all but P5), they reported becoming more aware of potential ethical issues when they started seeing "how good" the current level of technology is (e.g., *"I didn't think we'd be able to reproduce it to this level. We'd need technologies to counter the misuse of this technology."* (P1)). P3 shared that while developers are starting to engage in discussion related to these issues, at least in their social circles, people are not yet seriously considering these issues, nor have a clear idea or direction on what to do. Developers also felt that not enough actions are currently being taken. There were different opinions as to who needs to take the lead when it comes to implementing ethical and legal measures to counteract misuse. P3, for instance, viewed that "big players," such as large corporations investing in AI development, should play a bigger role. P4, on the contrary, stated that it will be dangerous for one party to decide how to ethically limit and/or control AI development, be it the government or big corporations, and emphasized that the society as a whole needs to engage in discussion and arrive at social consensus.

User comments showed more mixed opinions. Many were awed and moved by hearing the realistic AI voices, but they also shared fear, guilt, and discomfort caused by "uncanny valley" [8]. Compared to the developers' point of view, users had more pessimistic views on the technology, and a few questioned whether we should be developing such technology in the first place:

*"I became fearful as I was watching this [...] I feel lucky that I was born in an era when the AI isn't fully developed."*  
*"I think the technology is great, but I also wish we didn't overdo it. Do we really need to listen to songs created by machines? [...] there's no guarantee that the story in*

*movies won't come true where people who wanted support from AI eventually become controlled by them."*

These comments suggest that people's fears are influenced by how AI is portrayed in popular media. Several mentioned that the technology reminds them of science fiction TV shows or movies in which AI takes the role of the adversary. Compared to users, developers tended to be over-optimistic about the technology, trusting that counter technologies will exist and work well enough, with some passing the responsibilities to users to some extent (e.g., *"it is not the technology that is bad, but people who misuse it"* (P6), *"AI is just a tool"* (P3)). The discrepancy in how developers and users feel about the technology shows that it is important to encourage discussion on these issues between the two stakeholders, so the general public can be more informed about the actual state of the technology and the developers can understand how the technology is being received by the general public which will inevitably affect the future of such technology. P4 also emphasized the importance of communication with the general public about the technology as the developers' responsibility.

## 5.3 Meaning of "Creativity" in the Context of AI

A common question raised among developers and users concerns what it means to be "creative" and if AI generated voices can be considered as such. Defining what it means to be creative has ethical implications as determining who or what is responsible for the creative work affects the decision on who should "own" and benefit from the IP. Yet creativity has various definitions; one depends on the "ability to come up with ideas or artifacts that are new, surprising" [36] and another hinges on societal and cultural contexts and thus resists a concrete definition [37]. Some definitions disregard the notion of value and view it as irrelevant, instead emphasizing that creativity must look within the action of being creative itself [38].

As evident by the various perspectives, what is considered creative is widely contested, even more so once AI comes into frame. While it can certainly be agreed that creativity involves processes both cognitive and psychological, the question of whether AI can simulate these processes and the limitations to its approach via simulation remains [39]. P2 pointed out the lack of agency and intention from the AI, and questioned whether AI would be capable of creating something truly novel, stating: *"Even though AI can act perfectly like a human, we are just looking at the outcome. That doesn't mean the AI is really thinking about what it is expressing. It only learned from the data. So at the signal level, it might be similar but that is not an outcome from any kind of reasoning."* P4 also discussed how AI is good at interpolation from existing data sets, thus excelling in giving an "average" performance based on previous performances, but not something truly unique and novel. Some of the user comments also express similar sentiments, stating that the AI voice is "soulless", "lacks emotions", and sounds "too comfortable" or "too honest":

*"There isn't something deep or substantial, there's no emotion so I'm not moved [...] It'd sound more natu-*

*ral if the machine does adlibs or intentionally has trouble singing or barely sings in certain parts."*

*"The voice is the same but it's missing a soul [...] it's just following exactly what's written in the sheet music."*

Audry and Ippolito [40] propose a way of examining the relationship between AI and creativity by redirecting the focus from the artist, both human and artificial, towards the viewer. They draw on Foucault's designation of the "Meta-Artist" to postulate that regardless of whether or not AI can be creative/artists, they most certainly can give rise to an "artist function." As long as viewers continue to construct Meta-Artists, "artists will exist as social constructs" [40].

According to Gioti [39], while AI has generated impressive results in controlled environments, it has yet to break the barrier into autonomy. This leaves room for human artists and suggests the use of AI as an extended intelligence and thus another "actor contributing to a 'networked intelligence' that encompasses both humans and machines" [41]. From this, the concept of computational intelligence emerges where creative responsibility is shared among human and non-human actors where the latter is determined by the "extension of human intentionality through technological intentionality" [41]. The developers we interviewed also primarily considered the collaboration between humans and AI as a big area of opportunity, and none believed that the AI will truly replace the role of humans, partly because of the value ascribed to human skills [42]. Their perspective was focused on seeing AI's role as extending human abilities, similar to how Autotune is commonly used to manipulate singers' voices (P2). While this synergy between AI and humans is posed as ideal, it is limited in its exploration of ethical concerns and considerations of how such a teetering asymmetric relationship will affect both humans and machines.

#### 5.4 Human Rights, IP, and Other Legal Issues

Questions of who should make decisions about the ethics, legitimacy, and legality of SVS remain challenging and unanswered [43]. Sturm et al. [2] asks if the "lack of copyright protection of AI-generated results [is] adequate from a policy point of view" and recommends further "legal and socio-economic analysis." As of 2019, only the "UK, South Africa, Hong Kong, India, Ireland, and New Zealand have envisaged protection for computer-generated works granted to the person by whom the arrangements necessary for the creation of the work have been undertaken" [2]. Users and developers alike question who should be given the rights to the final product in situations where SVS is used to generate new voices using an artist's voice. P5 shared how they wished *"there are some rules about where the profits go to. Currently there are no laws about that."*

Regarding neighboring rights, P4 questioned how the financial gain should be distributed between the AI developer and the artists. Another question was about generating a deceased artist's voice – who should decide how the voice is created and used? P3 shared that family or people close to the deceased artists currently have the rights. In contrast, P6 brought up that the family members are still

not the artists themselves. Is it truly acceptable for the third party to make this decision? What if the voice sample is used with other samples to generate a new voice, so the voice data provider does not financially benefit from the result? Users commonly expressed concerns over the worth of human musicians in the event that AI becomes the norm rather than a collaborative tool. Both developers and users expressed conflicted feelings about SVS, with users expressing more negative sentiment such as fear or guilt, using phrases like 'superfluous man/잉여 인간'(i.e., humans with no roles or uses in the society). One proposed solution is to reintroduce scarcity via a Natural Talent certification, which identifies a composition or song as authentically made by human and differentiates it from music produced by AI [19]. However, the implementation is complicated given how human-computer collaboration on music already exists and is pervasive throughout the creation process.

In contrast to end users' concerns, developers are less worried about the "domination" of AI [44]. P2 stated that the individual performers should be able to maintain the performer's rights rather than the entertainment company they belong to, and that any speculated AI programs that will replace the artists will not succeed because of the backlash from the general public; and that coexisting is more likely to lead to success. Ultimately, P2 and P6 did not believe that any outcomes from AI can truly replace music created by humans because people will not want that to happen. As Sturm et al. state "humans still have an important involvement in creating music, even if assisted by an AI system" [2].

#### 6. CONCLUSION AND FUTURE WORK

This study offers initial insights into the perceptions of users and developers regarding ethical issues to consider when developing and implementing SVS technologies. Our findings highlight the discrepancies between user and developer perspectives regarding their envisioned use scenarios and attitudes, but also show that both stakeholders are similarly questioning creativity in the age of AI and concerned about human rights and IP issues.

This is a qualitative exploratory study with limited user data with a goal of enriching our understanding of the topic, not claiming a generalization of the findings. Future research involving more and varied developers and artists should be conducted to gain a more holistic understanding of the different stakeholders' viewpoints. In addition, the 3,075 user comments were predominantly from one culture and one platform and other cultures or platform users will have different perspectives, warranting further investigation. As this study analyzes user perception of the application of AI technologies presented in TV programs, it could have been influenced by how the usage was showcased in the media. However this is also realistically how the user perception is formed on new technology as media plays a significant role in our society. In the future, we also plan to conduct a follow-up study focusing on the perception of ethical issues from the artists' point of view.

## 7. ACKNOWLEDGEMENTS

We thank the interviewees who shared their honest thoughts and opinions about SVS development.

## 8. REFERENCES

- [1] Y. Goodfellow, I. Bengio and A. Courville, *Deep Learning*. Montréal, Canada: The MIT Press, 2018.
- [2] B. L. Sturm, M. Iglesias, O. Ben-Tal, M. Miron, and E. Gómez, “Artificial intelligence and music: open questions of copyright law and engineering praxis,” *Arts*, vol. 8, no. 3, p. 115, 2019.
- [3] Y. Song, S. Dixon, and M. Pearce, “A survey of music recommendation systems and future perspectives,” in *9th International Symposium on Computer Music Modeling and Retrieval*, vol. 4. Citeseer, 2012, pp. 395–410.
- [4] A. Van Den Oord, S. Dieleman, and B. Schrauwen, “Deep content-based music recommendation,” in *Neural Information Processing Systems Conference (NIPS 2013)*, vol. 26. Neural Information Processing Systems Foundation (NIPS), 2013.
- [5] C.-Z. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, I. Simon, C. Hawthorne, A. M. Dai, M. D. Hoffman, M. Dinculescu, and D. Eck, “Music transformer,” *arXiv preprint arXiv:1809.04281*, 2018.
- [6] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Sutskever, “Jukebox: A generative model for music,” *arXiv preprint arXiv:2005.00341*, 2020.
- [7] J. Engel, L. Hantrakul, C. Gu, and A. Roberts, “DDSP: Differentiable digital signal processing,” *arXiv preprint arXiv:2001.04643*, 2020.
- [8] M. Goto, T. Nakano, S. Kajita, Y. Matsusaka, S. Nakaoka, and K. Yokoi, “Vocalistener and vocawatcher: Imitating a human singer by using signal processing,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 5393–5396.
- [9] J. Lee, H.-S. Choi, C.-B. Jeon, J. Koo, and K. Lee, “Adversarially trained end-to-end korean singing voice synthesis system,” *arXiv preprint arXiv:1908.01919*, 2019.
- [10] S. Choi, W. Kim, S. Park, S. Yong, and J. Nam, “Korean singing voice synthesis based on auto-regressive boundary equilibrium gan,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 7234–7238.
- [11] P. R. Cook, “Singing voice synthesis: History, current work, and future directions,” *Computer Music Journal*, vol. 20, no. 3, pp. 38–46, 1996.
- [12] S. Oore, I. Simon, S. Dieleman, D. Eck, and K. Simonyan, “This time with feeling: Learning expressive musical performance,” *Neural Computing and Applications*, vol. 32, no. 4, pp. 955–967, 2020.
- [13] C.-Z. A. Huang, T. Cooijmans, A. Roberts, A. Courville, and D. Eck, “Counterpoint by convolution,” *arXiv preprint arXiv:1903.07227*, 2019.
- [14] N. Collins, “Trading faures: Virtual musicians and machine ethics,” *Leonardo Music Journal*, pp. 35–39, 2011.
- [15] S. Yong and J. Nam, “Singing expression transfer from one voice to another for a given song,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 151–155.
- [16] T. Saitou, M. Goto, M. Unoki, and M. Akagi, “Speech-to-singing synthesis: Converting speaking voices to singing voices by controlling acoustic features unique to singing voices,” in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2007, pp. 215–218.
- [17] Y. Wang, R. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, Z. Chen, S. Bengio *et al.*, “Tacotron: Towards end-to-end speech synthesis,” *arXiv preprint arXiv:1703.10135*, 2017.
- [18] A. Gibiansky, S. Ö. Arik, G. F. Diamos, J. Miller, K. Peng, W. Ping, J. Raiman, and Y. Zhou, “Deep voice 2: Multi-speaker neural text-to-speech,” in *NIPS*, 2017.
- [19] W. P. Jacobson, “The robot’s record: Protecting the value of intellectual property in music when automation drives the marginal costs of music production to zero,” *Loy. LA Ent. L. Rev.*, vol. 32, p. 31, 2011.
- [20] M. Avdeeff, “Artificial intelligence & popular music: Skygge, flow machines, and the audio uncanny valley,” *Arts*, vol. 8, no. 4, p. 130, 2019.
- [21] G. Lima and M. Cha, “Descriptive AI ethics: Collecting and understanding the public opinion,” p. *arXiv:2101.05957*, 2021.
- [22] L. Floridi, “Translating principles into practices of digital ethics: Five risks of being unethical,” *Philosophy & Technology*, vol. 32, no. 2, pp. 185–93, 2019.
- [23] K. K. Kimppa and T. I. Saarni, “Right to one’s voice?” in *Proceedings of ETHICOMP 2008: Living, Working and Learning beyond Technology*. ETHICOMP, 2008, pp. 480–88.
- [24] K. M. Scott, S. Ashby, D. A. Braude, and M. P. Aylett, “Who owns your voice? ethically sourced voices for non-commercial TTS applications,” in *CUI ’19: Proceedings of the 1st International Conference on Conversational User Interfaces*. Conversational User Interfaces (CUI), 2019, pp. 1–3.

- [25] B. C. Stahl, "From computer ethics and the ethics of AI towards an ethics of digital ecosystems," *AI and Ethics*, vol. 2, no. 1, pp. 65–77, 2022.
- [26] M. Hickok, "Lessons learned from AI ethics principles for future actions," *AI and Ethics*, vol. 1, no. 1, pp. 41–47, 2021.
- [27] R. Huang, B. L. T. Sturm, and A. Holzapfel, "Decentering the west: East Asian philosophies and the ethics of applying artificial intelligence to music," in *Proc. of the 22nd Int. Society for Music Information Retrieval Conf.*, 2021, pp. 301–309.
- [28] B. Zhang and A. Dafoe. (2019) Artificial intelligence: American attitudes and trends. [Online]. Available: <https://doi.org/10.2139/ssrn.3312874>
- [29] J. Dang and L. Li, "Robots are friends as well as foes: Ambivalent attitudes toward mindful and mindless ai robots in the United States and China," *Computers in Human Behavior*, vol. 115, 2021.
- [30] L.-M. Neudert, A. Knuutila, and P. N. Howard. (2020) Global attitudes towards AI, machine learning automated decision making. [Online]. Available: <https://oxcaigg.oii.ox.ac.uk/wp-content/uploads/sites/124/2020/10/GlobalAttitudesTowardsAIMachineLearning2020.pdf>
- [31] M. Clancy. (2021) Reflections on the financial and ethical implications of music generated by artificial intelligence. [Online]. Available: <http://www.tara.tcd.ie/handle/2262/94880>
- [32] N. Dell, V. Vaidyanathan, I. Medhi, E. Cutrell, and W. Thies, "'Yours is better!' participant response bias in HCI," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 1321–1330.
- [33] J. Corbin and A. Strauss, "Basics of qualitative research (4th ed.)," 2015.
- [34] C. E. Hill, B. J. Thompson, and E. N. Williams, "A guide to conducting consensual qualitative research," *The counseling psychologist*, vol. 25, no. 4, pp. 517–572, 1997.
- [35] A. Strauss and J. Corbin, "Basics of qualitative research 4th edition. technique and procedures for developing grounded theory," 2015.
- [36] M. A. Boden *et al.*, *The creative mind: Myths and mechanisms*. Psychology Press, 2004.
- [37] M. A. Boden, *Creativity and art: Three roads to surprise*. Oxford University Press, 2010.
- [38] A. Dorin and K. B. Korb, "Creativity refined: Bypassing the gatekeepers of appropriateness and value," in *Computers and creativity*. Springer, 2012, pp. 339–360.
- [39] A.-M. Gioti, "From artificial to extended intelligence in music composition," *Organised Sound*, vol. 25, no. 1, pp. 25–32, 2020.
- [40] S. Audry and J. Ippolito, "Can artificial intelligence make art without artists? ask the viewer," in *Arts*, vol. 8, no. 1. Multidisciplinary Digital Publishing Institute, 2019, p. 35.
- [41] J. Ito, "Extended intelligence," *PubPub*, 2016.
- [42] T. J. Pinch and K. Bijsterveld, "'Should one applaud?' breaches and boundaries in the reception of new technology in music," *Technology and Culture*, vol. 44, no. 3, pp. 536–59, 2003.
- [43] N. Mirra, "Putting words in your mouth: The evidentiary impact of emerging voice editing software," *Richmond Journal of Law and Technology*, vol. 1, 2018.
- [44] G. Velarde. (2021) Artificial intelligence trends and future scenarios: Relations between statistics and opinions. [Online]. Available: <https://www.computer.org/csdl/10.1109/CogMI52975.2021.00017>