THE IMPACT OF AI IN THE FIELD OF SOUND FOR PICTURE.
A HISTORICAL, PRACTICAL, AND ETHICAL CONSIDERATION

DAN-ȘTEFAN RUCĂREANU
I.L. Caragiale UNATC, Bucharest, Romania
dan.rucareanu@unatc.ro

LAURA LĂZĂRESCU-THOIS
I.L. Caragiale UNATC, Bucharest, Romania.
University Politehnica of Bucharest
laura.lazarescu@unatc.ro

Abstract: The paper covers the emergence and evolution of AI in the commercial field of audio: from the pioneering German audio software company Prosoniq Products Software, which first used artificial neural networks for commercial audio processing, creating in 1997 the program Pandora Music Decomposition Series that managed to separate music into its components, to the indispensable technologies for sound processing involving machine learning (iZotope and Audionamix ADX Trax Pro 3), to the spleeter technology, ChatGPT from OpenAI that applies AI technology also to the area of sound processing, to Adobe or ElevenLabs that offers online AI services with Text to Speech and Speech to Speech functions.
The second part of the paper will review specific audio plug-ins that use Artificial Intelligence, making it possible to identify to what extent AI is an important tool in the field of sound and music, but also what are its current limitations.
The last part of the paper refers to the ethics of AI in the field of sound. Even if plug-ins facilitate the work process in sound design, music production, voice cloning and synthesis, or sound restoration, thus reducing costs and personnel, what are the ethical limits of using these tools?

Keywords: AI, sound restoration, generative speech, generative music, spleeter, machine learning, demixing, demucs.

How to cite: Rucăreanu, D-Şt. and Lăzărescu-Thois, L. (2024) "The Impact of AI in the Field of Sound for Picture. A Historical, Practical, and Ethical Consideration", *Concept* 2(29), pp. 112-128. DOI: https://doi.org/10.37130/bf780v59

Introduction

The emergence of artificial intelligence, innovations, and technological advances in various fields, including sound, makes this a fascinating topic but also raises several questions and fears about the future of jobs in the field of sound, sound production and post-production, and sound workflow. On the internet, one can find a multitude of applications and programs, with the help of which it is possible to create in just a few seconds various voiceovers, clips for social networks, different styles of songs with their lyrics and themes, and even sound effects for video or movie scenes starting from just a few keywords.

To know where we are and where we are going, it is important to understand the starting point of AI in sound. A brief history showing the most significant technological advances and developments helps us anticipate what is to come. It is good to note that this history is based on two divergent but complementary objectives: one seeks to generate audio content, and another wants to separate the elements that make up the audio signal. Both are a precursor step to consolidating an artificial general intelligence that has the human-like ability to separate and distinguish audio elements from a sound mix (e.g. cocktail party effect) and

generate or imagine sound effects, music, and voices which are validated by human perception.

If the generated audio content is already in the public eye, especially through generative AI models of music and voice, and lately also of sound effects, in terms of audio demixing, the AI revolution is happening behind the scenes, but just as effervescently.

A brief history of AI in audio

Around 1997, the German audio software company Prosoniq Products Software pioneered the use of artificial neural networks for commercial audio processing, creating a program called Pandora Music Decomposition Series that uses an ANN (artificial neural network)-based algorithm. It was designed to split a complete composite music track into its basic musical components (voice, instruments, and residual sound). "This is done using Prosoniq's proprietary Virtual Hearing technology […] based on a new approach to modeling the human ear and auditory system, and a signal representation developed using recent discoveries in the fields of neural network processing, signal decomposition, and source separation" (Computer Music Journal, 1997, p. 116). A year later, Prosoniq released another ANN-based product called SonicWorX Artist that could isolate the voice from the audio background, be it noise, music, or even reverberation, with a relatively low percentage of post-processing artifacts (Lehrman, 1998).

The algorithm served as the foundation for other innovations that followed: in 2003, Hartmann created a keyboard synthesizer called Neuron that used the same algorithm (Reid, 2003), and in 2010, Prosoniq released VuvuX to remove the noise of vuvuzela from the stadium ambiance of the 2010's FIFA Football World Cup in real time (Sound on Sound, 2010). In 2012, Zynaptiq bought Prosoniq's technology and created the plug-in software called Unveil that could reduce or even eliminate reverberation and echo from the audio signal, with minimal degradation of the useful signal (Zynaptiq, 2013). If until then, the recording of a natural sound with a good signal-to-noise ratio, without an excess of reverberation, were mandatory conditions for a quality recording, from that moment on, even a heavily reverberated sound could be usable.

The iZotope suite, launched in 2001, has become indispensable in audio production, be it audio restoration, processing, or mastering, quickly becoming a standard in music production and film and media postproduction. Since 2016, iZotope RX has been using machine learning to introduce advanced audio processing modules such as Dialog Isolate, the Repair Assistant module, and De-Reverb, all based on complex algorithms (Wichern, 2017; Rose, 2018).

Founded in 2008, Audionamix has been improving the technology to separate stems from music or isolate a voice – a path-breaking innovation that finally, in 2016, led to the ADX Trax Pro program that used deep learning technology with which it was not only possible to separate voices from instruments, but even to exclude all melodic content from a file, allowing the creation of new mixes, arrangements, upmixes from mono to stereo or surround (Thomas, 2017; Rose, 2018).

2016 marked many advances in voice generation and processing: it was the year Adobe publicly presented a Text to Speech prototype called VoCo, nicknamed "Photoshop for audio", with which a voice could be cloned, and then one could modify the speech and add other words or phrases, in the same voice (Adobe, 2016). Although revolutionary, from the moment of the announcement, VoCo was quickly "buried", mainly due to ethical concerns that were voiced in the public space. It was obvious that neither society nor the law was yet ready to accept the technology presented in the public domain by Adobe. However, in the same year, Google's DeepMind presented the WaveNet project, a Text to Speech synthesizer that was able to reproduce text with artificial voice (van den Oord and Dieleman, 2016). It was much easier to accept because the voice wasn't as natural as Adobe's and similar predecessor technologies were already acknowledged by the public.

Interest in technological advances in this field was maintained in the following years. In 2019, open-source demixing technologies were developed through Spleeter (Moussallam, 2019) and Demucs (Défossez et al, 2019). Thus evolved the ability to separate different instruments and voices from a piece of music. Virtually all current demixing third party plug-ins use a modified version of these deep machine-learning models. In 2020, a sneak peek of the documentary "The Beatles: Get Back" appeared on YouTube (The Beatles, 2020), which revealed archive footage processed using Spleeter technology in combination with other programs to separate the band's recording sessions into stems (Hurwitz, 2021).

Also in 2020, an advertisement for the program Descript promised to convert the recorded voice of a speaker into text (Speech to Text) and then allow it to be edited and rewritten (regenerating the voice from the text with the same vocal timbre). Descript could also detect and remove so-called "filler words" (e.g. 'uhm', 'uh') from the recording, replace mistakes or stutters with other words having the same vocal timbre, or transform a recording with its own voice (not another person's voice, for ethical reasons), having a different state, mood (Descript, 2020).

The year 2022 marked the launch of OpenAI's ChatGPT, which brought AI technology to the fore with all its modules, including the one for audio (OpenAI, 2024a).

The year 2023, came with an avalanche of innovations: Adobe introduced Podcast AI, a program that transformed the voice from a non-professional recording to one that sounds like a studio recording (Adobe, 2023a), then teased its followers with the presentation of a yet unreleased program called Adobe DubDubDub that permits dubbing images with a voice in multiple languages (Adobe, 2023b). Thus, the dubbing of games, movies, and VR projects could become accessible regardless of geographical area and language barriers. Proof: the 70 languages and 140 dialects available.

ElevenLabs offers online AI services with Text to Speech and Speech to Speech features, becoming one of the top-rated AI firms due to the quality of the sound delivered. Currently, you can offer your voice in the Voice Library and receive various benefits, including financial ones, when it is used (ElevenLabs, 2024).

Also in 2023, other online AI programs appeared that allow changing the voice, even in real-time, with a cloned voice (including that of a public figure) and, of course, changing one singer with another on a music negative.

Accordingly, the first commercial song composed exclusively with AI was released: "Heart on My Sleeve" by ghostwriter977, a TikTok user, with AI-generated vocals by Drake and The Weekend. The song was subsequently removed from all streaming platforms for copyright claims formulated by Universal Music Group (Paul and Millman, 2023). Also in 2023, Google launched MusicLM, which enables the Text to Music module. One can create music starting not only from a simple text but also from the name of a painter and the title of his painting (Agostinelli et al, 2023).

The Meta company presented VoiceBox for cloning and voice generation by text (Meta, 2023a), and at the end of 2023 proposed a prototype project called Audiobox through which users can create voices, music, and sound effects through text. In the case of the voice, they can determine how it sounds, choosing the intonation, but also the location from which it is heard (Hus et al, 2023). Also in 2023, Meta released an open-source AI tool for music and sound effects named AudioCraft that can generate music and sound effects based on textually guided generation (Meta, 2023b).

If in 2000, Google could clone the voice after only five seconds of listening, in 2023, Microsoft reduced cloning to three seconds of listening, and Meta Voicebox reduced it further to two seconds (Meta, 2023c).

In 2024, OpenAI created Sora, which generates extremely realistic video images from text, featuring landscapes, multiple characters, and complex camera movements (OpenAI, 2024b). Immediately, ElevenLabs created the sound for the images generated by Sora with a Text to Sound model that was released as Sound Effects with limited beta access for testing until June 2024, when it will become

public. In addition, ElevenLabs also offers an AI Dubbing program (Staniszewski, 2024), similar to the still unreleased Adobe DubDubDub.

A couple of months later, new Text to Sound effects AI models surfaced (e.g. Stableaudio, Myedit, OptimizerAi, Plugger AI, Lovo AI, etc.), strengthening the number of still limited quality tools for generating natural sound effects in addition to voice and music. Also, in May 2024, Google proposed a counterpart to Sora AI, named Veo, showing that this race has only just begun (Collins and Eck, 2024).

But video without sound remains incomplete, so the next natural step was researching the generation of purposeful sound to silent pictures with AI models. One breakthrough was announced in 2016 by MIT Computer Science and Artificial Intelligence Lab in collaboration with Google, which revealed how AI could produce realistic sounds that were in sync with the moving images and could fool humans into thinking that they were watching a real sync recording (Owens et al, 2016). This research has two major implications: one is the ability to predict and synthesize natural synchronous sound to images, and the other can help robots better understand how objects interact with the environment. Adobe followed shortly with its own research that could generate sound clips to match video (Zhou et al, 2018).

In 2020, a new research paper for a multilayered machine-learning program named AutoFoley was released in IEEE Transactions on Multimedia (Prevost and Ghose, 2020). This multi-modal AI can analyze a movie clip and create realistic sound effects (named Foley by the cinema community) that sync to the image. As is the case with previous models, the sync is more an approximation than a perfect match. Still, in a blind survey made to compare original to generated foley to movie clips, more than 60% of the respondents opted for the generated sound effects as being the original version (Prevost and Ghose, 2020).

One year later, Runaway AI released Soundify, a prototype AI model that can concomitantly generate sound effects and sound ambiances in sync with the picture (Lin et al, 2021). By its third iteration, Soundify could localize sound in the image by using panning and volume (Lin et al, 2023). In 2024, Pika Labs released a generative sound effect to video AI model that is open for all registered users to test and use (Morrison, 2024).

In the music domain, SunoAI allows the generation of music (instruments, vocals, and lyrics), based only on a description entered by the user via text (Hiatt, 2024). SunoAI partnered with Microsoft Copilot to further develop the technology (Microsoft Copilot, 2023). UdioAI, made by a former Google DeepMind team, joined in as a rival to SunoAI and the quality of its generated music attracted the attention of several artists in the music industry (Nuñez, 2024). Following the

lead of SunoAI and UdioAI, ElevenLabs (Music), Google (Music), and other AI models are now competing to take the lead in the generative music domain (taking into account all ethical concerns that have already surfaced in the industry).

New programs are constantly being developed, algorithms are being improved, and technology is being perfected. There is also a "Crowdsourcing AI to Solve Real-World Problems [that] enables data science experts and enthusiasts to collaboratively solve real-world problems, through challenges" (Aicrowd, 2024), which includes each year sound related challenges, whose results – research papers and source code – are open to the public.

The number of generative audio AI models increases each month and in the domain of neural network sound processing, a comparison of the latest commercially available programs that use AI has appeared online (Devideconcept, 2024), demonstrating the advancement and current limitations of unmixing and restoration AI technologies.

AI in practice

The AI revolution in audio is based on two distinct and complementary pillars. Both technologies aim to reproduce the psycho-acoustic perception and determine the cognitive-creative capacity of the human being in the audio field. It thus provides tools that facilitate and even stimulate the creators' artistic process and develops towards a complement of what may eventually be called full or general artificial intelligence, which wants to join creators in their artistic endeavors.

The considerations surrounding this vision yield a comprehensive spectrum of interpretations and come with many concerns from the film and media sound industry. Without prior research into what new AI can deliver, now and in the future, speculation takes the place of insight. But the process turns out to be arduous and not without problems, especially if we talk about the applicability of these AI models in the film and media sound workflow, where the reluctance to embrace the new remains a syndrome inherited from the time of the digital revolution. We observe that many of the new AI technologies still need human input, both perceptual–determinative (needing the ability to perceive all the possibilities and to determine what is the best choice esthetically and/or functionally) and creative–organizational (ultimately, AI models can remix what they have learned but cannot create new content that is unique and also organized, with narrative meaning), even if they eventually provide results that could not be achieved otherwise.

In the following, we will present two examples of audio-visual products, whose sound we have processed with emerging AI technologies. Based on them,

we can form an overview of the possibilities offered by these tools and discover what their current limits are.

The first example refers to a segment of a short student film from the archive of the "I.L. Caragiale" UNATC (*People are not Goats*, 1971; Arhiva Activă UNATC, 2024), and illustrates two stages of the research: firstly, the integration of AI tools in the workflow of restoring the old film sound and secondly, the process of dubbing the original language in English, with the original actors' voices. The second example refers to a sound design for a short video clip, made only using AI-generated sound effects with minimal intervention in editing and signal processing, to see if it can resemble a sound design made with natural sound effects.

The idea of using AI tools to process sound as efficiently as possible is not a new one, as evidenced by the fact that in the last year, there have been many AI-based plug-in programs that allow polishing the sound or cleaning it in unique ways, that were quickly adopted by the sound industry in the field of film and media. Thus, integrating AI technology into the sound restoration of short student films from the UNATC archive proved to be a natural choice.

These AI tools do not work without user input, and in most cases, given that they have minimal modifiable parameters, they can produce results with many artifacts if not set precisely. In other words, it is the user who decides, through his perceptive-auditory skill, what is the best result that the program can create.

For the experiment of restoring an old film sound with the help of AI tools, we have used Steinberg Nuendo 13 as the DAW and AI-based plug-ins and programs from companies such as Steinberg (VoiceSeparator, Spectral Layers Pro 10), Accentize (dxRevive, Spectral Balance, Chameleon), Supertone (Clear), iZotope (Rx Advance 10) and Waves (Clarity Vx). Starting from a mono source with parasitic noises and distortion, the AI programs managed to help in the process of cleaning and restoring the sound, keeping the sound quality and the original information of the recording, separating the voice, the music, and the sound effects in different stems to be processed separately, and removing the artifacts produced by the passage of time.

In the result which can be viewed online (Rucăreanu, 2024a), we compared the original version of the old film sound (recorded directly from the film's negative), the conventionally processed version – without the help of AI (made by sound designer Ioan Bain), and then the AI-restored version (our version).

Following the good results from the restoration stage, we set out to explore other ways we could use AI technology. Thus, we decided to continue the experiment into a second stage and used the voice cloning technology to

dub the film in English, keeping the authenticity of the actors' voices. We chose to use the automatic dubbing module from ElevenLabs (auto-translation, auto-voice cloning, and auto-dubbing with sync to picture), which had just appeared at the beginning of 2024, but unfortunately, the result did not satisfy us (Rucăreanu, 2024b). Then, we opted to manually clone and duplicate each actor and each line of text, also using the voice synthesis provided by the company ElevenLabs (Text to Speech). After several attempts to generate a voice which would transmit emotion, we managed to get quite close to the original Romanian version, in a relatively short time. Note that in this process, we only used AI synthesis and synchronized the audio fragments manually (Rucăreanu, 2024b).

Figure 1. Steinberg Nuendo 13 DAW interface with some of the AI-based plugins.

For the second experiment, we aimed to create a sound design for a specific video, using only AI-generated sound effects. This task was inspired by a contest found on the website asoundeffects.com (Andersen, 2024a) where the participants had to design a sound for a short clip using only recorded content or only synthesized content, plug-in alterations being allowed in both cases, as long as the sound design was deemed suitable for the given image (ASoundEffect, 2024). Our goal was to see if we could create a sound design using only AI-generated (synthesized) sound effects that still seemed created with only (natural) recorded sound effects.

Earlier this year, both Meta and ElevenLabs released their Text to Sound Effects AI models that we got test access to. Meta's Text to Speech AI model is named *Audiobox* and ElevenLabs's Text to Sound Effects AI mode is named *Sound Effects*. The technology is revolutionary and unique, never released before for the public, at least at this level of sound quality, where it manages to generate sound effects closely resembling naturally recorded sounds. We thought it was an opportune time to test the capability of these models and see if they could create a convincing sound design based on a detailed text prompt.

For both AI models, we used the same detailed text prompt that should have generated an appropriate sound design for the short clip provided in the contest. As can be seen (and heard) online, both Meta's Audiobox result (Rucăreanu, 2024c) and the ElevenLabs' Sound Effects result (Rucăreanu, 2024d) were not very convincing.

Although both programs generated different results, we quickly realized that neither of them could (yet) build a complex narrative sound structure. At least, not for what we needed. At the current prototype level, both AI models can only generate single sound effects, although Meta states that it is possible to request a sound "phrase". So, we set out to handcraft sound design while still using AI audio generation. We opted for ElevenLabs because the result seemed more qualitative and natural.

We proposed a stricter canon: to not exceed 20 generated sound effects and to create a sound design in which we do not use any processing that could alter the sound quality of the generated sound effects, other than the regular mixing plug-ins (equalizer, compressor, limiter). Due to limited control of the generated sound effects, we allowed ourselves to change the speed and pitch of a couple of generated sounds, and also applied a distortion for a single generated sound, as an amplification effect, but except this, we did not use any sound quality alteration, so as to keep the AI-generated sounds in their original state as much as possible. The generated sound effects were synched and edited to the picture in the same way as we would have done if we had only recorded sound effects. All sound effects were generated directly on the company's website, and for each written prompt the model offered us five variants to choose from and download locally.

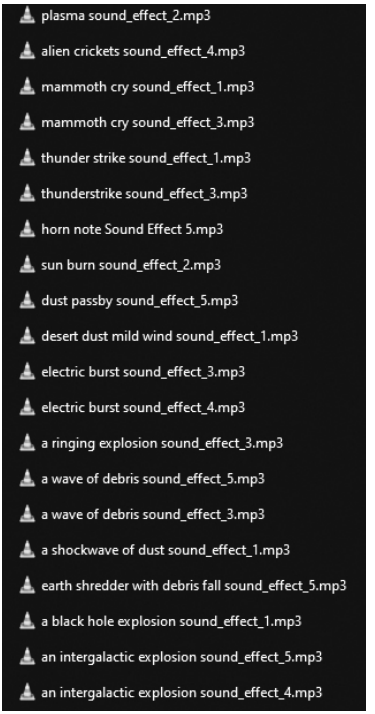

Figure 2. Chosen AI-generated sound effects

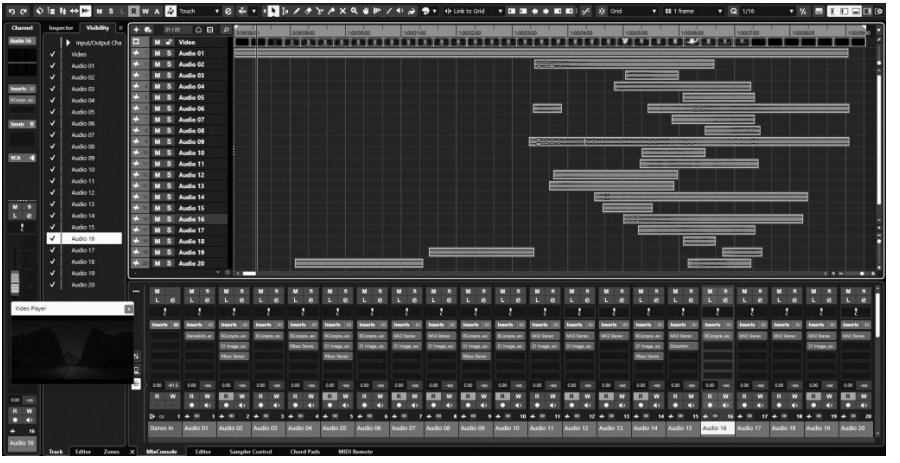

Figure 3. Sound design project created with AI-generated sound effects in ElevenLabs Sound Effects

The result appears to be impressive given the fact that we have not used a single recorded sound for this project. The result can be seen and heard online (Rucăreanu, 2024e) and can be compared with the winning projects which used strictly recorded sound effects (Andersen, 2024b). However, it must be considered that the user input was a key factor in successfully completing the research and acknowledge that the AI model used was still in its early stage of development and beta test.

Following the examples given in this article, we can state that the advancement of the technology of AI models in the field of audio for picture is no longer at its early stage, but still has a long way to go until it can be used commercially and integrated into the workflow of sound design for movies and media.

What remains questionable is the possible future capacity of AI models to perform in the field of sound for film and media without needing user input and thus replace the conventional job of sound designers, composers, and actors, who we have been used to for decades. It may seem that this 'reality' is still far away but given the recent fulminant (r)evolution of AI, anything is possible in the future.

Ethics in AI

The association between ethics, a branch of philosophy dating back thousands of years, from Ancient Greece, and artificial intelligence, a modern human discovery, is an interesting topic. Technology ethics is usually categorized as a branch of applied ethics, and AI ethics has its roots in computer ethics.

The accessibility of AI tools brings with it a series of questions about how we can properly use AI technology. Although the advancement of technology goes far beyond devising a solid legislative framework regarding AI rules, limits, and prohibitions, large companies have established a set of ethical rules and standards, either out of precaution or out of concern for the public opinion.

Thus, the company ElevenLabs launched in June 2023 the AI Speech Classifier with the help of which anyone can upload audio samples and find out whether they were created by AI through ElevenLabs or not. At the same time, they are constantly working on regulations regarding the use of certain words (defamatory, violent, fraudulent, abusive), but also in the creation of fake news videos with a political theme, to manipulate voters. To stop this, "no-go voices" were created, with the help of which it is possible to block the creation of voices that are similar to those of candidates in various elections in the US or UK, and the technology will be developed with other languages, in other countries.

Adobe is constantly developing both automated and human verification techniques to prevent the misuse of technology, misinformation, and stereotypes,

and to create the best products, focusing on ethical impact, but not slowing down innovation.

Sora AI limits the generation of videos with themes of violence, sexuality, and even celebrity likeness, and is working on improved methods to detect whether a video was generated by their program or not.

Meta talks about the five "pillars" of responsible use of AI: privacy and security, fairness and inclusion, robustness and safety, transparency and control, and finally, accountability and governance.

Google explains a whole series of goals that they pursue through AI, including the principles of non-discrimination, combating bias, stereotypes, differences in gender, religion, sexual orientation, etc., which work both ways, maintaining a balance between extremes (see the situation with Google GeminiAI that created black Nazis in order not to discriminate – Grant, 2024), but also areas in which they will not implement the technologies if they attack human rights or international regulations.

OpenAI talks about the development of safe and responsible AI, while still supporting the idea of opening the source code of every created and released model to the public in order not to allow any one entity to have a monopoly on the future.

As society understands the potential of AI, it will react, adapt, and create a safe environment for the use of this new technology that will eventually revolutionize the lives of each of us, just as electricity, industrialization, and digitalization did.

Speaking of the film and media industry, a first reaction was observed among the Screenwriters and Actors Union of Hollywood which ended with a set of rules that should primarily protect the interests of creators. Other creative professions, including sound engineers and designers, have launched discussions in order to explore the possibilities of keeping as much as possible of what already works in the face of a technology that seems to aim to revolutionize everything.

At the international level, the European Union voted on a law that regulates the use of AI technology at all levels. The law has become a global benchmark, being taken up by the US and other states.

Conclusion

We are left with many unanswered questions: Are all these regulations and standards enough to protect the interests of humanity? Are art creators protected in the face of an expansive AI that seems to capture and facilitate every aspect of the artistic act? Is the organic aspect of a sound design still preserved in the face of a technology that ends up synthesizing natural recordings (existing or imagined)? Will the jobs of sound engineers, sound designers, composers, singers, and actors still exist in the next quarter of a century? Will AI technology remain a

certainty in the future (as digital technology was) or will it be replaced by another, more suitable one?

These are questions for which we can imagine an endless number of possibilities, each with its own valid and supported argument. But in the end, as with the humanistic revolution, the industrial revolution, or the digital revolution, the one who tries to understand the change is the one who can adapt to it. Revolution does not necessarily imply a replacement of fundamental values, but a special look at them. Finally, the arts cannot disappear: after all, they prevailed before AI, in the face of previous revolutions. But more likely, artists will have to adapt to new ways of relating to reality, using what AI will make available to them. Now and in the future.

We cannot speculate what will happen many years from now, but in the near future we can say with certainty that it will not be AI technology that will replace the sound designer, but rather that the sound designer who understands and uses AI will replace the "sound designer".

Perhaps in five, ten, or fifteen years, a sound designer will no longer be called "the one who knows how to produce and build narrative structures of sounds", but "that artist who is able, through language and emotion, to collaborate with and to guide his AI partner", to create artistic products that will be accepted by the public of those times.

Online references:

1. Adobe (2016) *#VoCo. Adobe Audio Manipulator Sneak Peak with Jordan Peele* [Online]. Available at: https://www.youtube.com/watch?v=I3l4XLZ59iw/ (Accessed: 20 May 2024).
2. Adobe (2023a) *Behind the Tech: Enhance Speech in Adobe Podcast* [Online]. Available at: https://research.adobe.com/news/behind-the-tech-enhance-speech-in-adobe-podcast/ (Accessed: 20 May, 2024).
3. Adobe (2023b) *#ProjectDubDubDub | Adobe MAX Sneaks 2023* [Online]. Available at: https://www.youtube.com/watch?v=fZY-Cv1Q8NY/ (Accessed: 20 May 2024).
4. Agostinelli, A. et al (2023) *'MusicLM: Generating Music From Text'*, arXiv:2301.11325v1 [cs.SD] [Online]. Available at: https://arxiv.org/pdf/2301.11325 (Accessed: 20 May 2024).
5. Aicrowd (2024) *Crowdsourcing AI to Solve Real-World Problems* [Online]. https://www.aicrowd.com (Accessed: 20 May 2024).
6. Andersen, A. (2024a) *Huge sound design contest! 7 chances to win fantastic prizes from Mattia Cellotto* [Online]. Available at: https://www.asoundeffect.com/sounddesign24/ (Accessed: 20 May 2024).
7. Andersen, A. (2024b) *Here are the winners of the Mattia Cellotto sound design contest!* [Online]. Available at: https://www.asoundeffect.com/sounddesign-contest-results/ (Accessed: 20 May 2024).
8. Arhiva Activă UNATC (2024) *Oamenii nu sunt capre* [Online]. Available at: https://arhiva.unatc.ro/filme/oamenii-nu-sunt-capre/ (Accessed: 20 May 2024).
9. ASoundEffect (2024) *Sound Design Contest: Create the sound for this video for 7 chances to win wild prizes!* [Online]. Available at: https://www.youtube.com/watch?v=X_wsXpUxOaY (Accessed: 20 May 2024).

10. Collins, E. and Eck, D. (2024) *New generative media models and tools, built with and for creators* [Online]. Available at: https://blog.google/technology/ai/google-generative-ai-veo-imagen-3/ (Accessed: 20 May, 2024).
11. Computer Music Journal (1997) *'Products of Interest'*, *Computer Music Journal*, vol. 21, no. 3, p. 116. [Online]. Available at: http://www.jstor.org/stable/3681029 (Accessed: 20 May 2024).
12. Défossez, A. et al (2019) *'Music Source Separation in the Waveform Domain'*, arXiv:1911.13254v1 [cs.SD] [Online]. Available at: https://arxiv.org/pdf/1911.13254v1 (Accessed: 20 May 2024).
13. Descript (2020) *Introducing Descript* [Online]. Available at: https://www.youtube.com/watch?v=Bl9wqNe5J8U/ (Accessed: 20 May 2024).
14. Devideconcept (2024) *Audio AI Comparisons* [Online]. Available at: https://divideconcept.github.io (Accessed: 20 May 2024).
15. ElevenLabs (2024) *Introducing Voice Actor Payouts* [Online]. Available at: https://elevenlabs.io/blog/introducing-voice-actor-payouts/ (Accessed: 20 May 2024).
16. Grant, N. (2024) *Google Chatbot's A.I. Images Put People of Color in Nazi-Era Uniforms* [Online]. Available at: https://www.nytimes.com/2024/02/22/technology/google-gemini-german-uniforms.html (Accessed: 20 May 2024).
17. Hiatt, Brian (2024) *A ChatGPT for Music Is Here. Inside Suno, the Startup Changing Everything* [Online]. Available at: https://www.rollingstone.com/music/music-features/suno-ai-chatgpt-for-music-1234982307/ (Accessed: 20 May 2024).
18. Hsu et al (2023) *Audiobox: Unified Audio Generation with Natural Language Prompts* [Online]. Available at: https://ai.meta.com/research/publications/audiobox-unified-audio-generation-with-natural-language-prompts/ (Accessed: 20 May 2024).
19. The Beatles (2020) *The Beatles: Get Back - A Sneak Peek from Peter Jackson* [Online]. Available at: https://www.youtube.com/watch?v=UocEGvQ10OE (Accessed: 20 May 2024).
20. Hurwitz, M. (2021) *The Making of The Beatles' Let It Be and Peter Jackson's Get Back Peter Jackson's The Beatles: Get Back* [Online]. Available at: https://www.soundandvision.com/content/making-beatles-let-it-be-and-peter-jacksons-get-back-peter-jacksons-beatles-get-back/ (Accessed: 20 May 2024).
21. Lehrman, P. D. (1998) *'Prosoniq Sonicworx Artist. Audio Editing Software.' Sound on Sound*, vol. 13, issue 5, March. [Online]. Available at: https://www.soundonsound.com/reviews/prosoniq-sonicworx-artist/ (Accessed 20 May 2024).
22. Lin, David Chuan-En et al (2021) *Soundify: Matching Sound Effects to Video*, arXiv:2112.09726v1 [cs.SD] [Online]. Available at: https://arxiv.org/pdf/2112.09726v1 (Accessed: 20 May 2024).
23. Lin, David Chuan-En et al (2021) *Soundify: Matching Sound Effects to Video*, arXiv:2112.09726v3 [cs.SD] [Online]. Available at: https://arxiv.org/pdf/2112.09726v3 (Accessed: 20 May 2024).
24. Meta (2023a) *Introducing Voicebox: The Most Versatile AI for Speech Generation* [Online]. Available at: https://about.fb.com/news/2023/06/introducing-voicebox-ai-for-speech-generation/ (Accessed: 20 May 2024).
25. Meta (2023b) *Open sourcing AudioCraft: Generative AI for audio made simple and available to all* [Online]. Available at: https://ai.meta.com/blog/audiocraft-musicgen-audiogen-encodec-generative-ai-audio/ (Accessed: 20 May 2024).
26. Meta (2023c) *Introducing Voicebox: The Most Versatile AI for Speech Generation* [Online]. Available at: https://about.fb.com/news/2023/06/introducing-voicebox-ai-for-speech-generation/ (Accessed: 20 May 2024).
27. Microsoft Copilot (2023) *Turn your ideas into songs with Suno on Microsoft Copilot* [Online]. Available at: https://www.microsoft.com/en-us/microsoft-copilot/blog/2023/12/19/turn-your-ideas-into-songs-with-suno-on-microsoft-copilot/ (Accessed: 20 May 2024).

28. Morrison, R. (2024) *Pika Labs sound effects now available for all users — I tried it and here is how it sounds* [Online]. Available at: https://www.tomsguide.com/ai/ai-image-video/pika-labs-sound-effects-now-available-for-all-users-i-tried-it-and-here-is-how-it-sounds/ (Accessed: 20 May 2024).
29. Moussallam, M. (2019) *Releasing Spleeter: Deezer Research source separation engine* [Online]. Available at: https://deezer.io/releasing-spleeter-deezer-r-d-source-separation-engine-2b88985e797e/ (Accessed: 20 May 2024).
30. Nuñez, Michael (2024) *Former Google DeepMind researchers launch AI-powered music creation app Udio* [Online]. Available at https://venturebeat.com/ai/former-google-deepmind-researchers-launch-ai-powered-music-creation-app-udio/ (Accessed: 20 May 2024).
31. OpenAI (2024a) *Navigating the Challenges and Opportunities of Synthetic Voices* [Online]. Available at: https://openai.com/index/navigating-the-challenges-and-opportunities-of-synthetic-voices/ (Accessed: 20 May 2024).
32. OpenAI (2024b) *Video generation models as world simulators* [Online]. Available at: https://openai.com/index/video-generation-models-as-world-simulators/ (Accessed: 20 May 2024).
33. Owens, A. et al (2016) *Visually Indicated Sounds*, arXiv:1512.08512v2 [cs.CV] [Online]. Available at: https://arxiv.org/pdf/1512.08512 (Accessed: 20 May, 2024).
34. Paul, L. and Millman, E. (2023) *Viral Drake and The Weeknd AI Collaboration Pulled From Apple, Spotify* [Online]. Available at: https://www.rollingstone.com/music/music-news/viral-drake-and-the-weeknd-collaboration-is-completely-ai-generated-1234716154/ (Accessed: 20 May 2024).
35. Prevost, John J. and Ghose, Sanchita (2020) *AutoFoley: Artificial Synthesis of Synchronized Sound Tracks for Silent Videos with Deep Learning*, arXiv:2002.10981v1 [cs.SD] [Online]. Available at: https://arxiv.org/pdf/2002.10981/ (Accessed: 20 May 2024).
36. Prosoniq (2010) *Prosoniq get the horn. Free World Cup audio plug-in released* [Online]. Available at: https://www.soundonsound.com/news/prosoniq-get-horn/ (Accessed: 20 May 2024).
37. Reid, G. (2003) '*Hartmann Neuron. Neuronal Resynthesizing Keyboard*' [Online]. Available at: https://www.soundonsound.com/reviews/hartmann-neuron/ (Accessed: 20 May 2024)
38. Rose, Jey (2018) Neural Networks: A new way to think about processing, CAS Quarterly [Online]. Available online at: https://digital.copcomm.com/i/933783-winter-2018/29? (Accessed: 20 May 2024).
39. Rucăreanu, D. Ș. (2024a) *AI Restored example - student short film from UNATC archives* [Online]. Available at: https://www.youtube.com/watch?v=5wc5HyzWbho (Accessed: 20 May 2024).
40. Rucăreanu, D. Ș. (2024b) *AI English Dubbing Example - student short film from UNATC archives* [Online]. Available at: https://www.youtube.com/watch?v=sKjcw1ADxlQ (Accessed: 20 May 2024).
41. Rucăreanu, D. Ș. (2024c) *Audio Example 1 - generated with Meta Audiobox (Beta)* [Online]. Available at: https://www.youtube.com/watch?v=lpas-3FkCKU (Accessed: 20 May 2024).
42. Rucăreanu, D. Ș. (2024d) *Audio Example 2 - Generated with ElevenLabs Sound Effect (Beta)* [Online]. Available at: https://www.youtube.com/watch?v=lmo099JKVhc (Accessed: 20 May 2024).
43. Rucăreanu, D. Ș. (2024e). *AI Sound Design Example - audio elements generated with ElevenLabs Sound effects (Beta)* [Online]. Available at: https://www.youtube.com/watch?v=teJ0g2pkmtI (Accessed: 20 May 2024).
44. Staniszewski, M. (2024) *Introducing Dubbing Studio. Localize videos with precise control over transcript, translation, timing, and more* [Online]. Available at: https://elevenlabs.io/blog/introducing-dubbing-studio/ (Accessed: 20 May 2024).

45. Thomas, B. (2017) *Audionamix ADX Trax Pro 3 SP. Speech & Melody Separation Software* [Online]. Available online at: https://www.soundonsound.com/reviews/audionamix-adx-trax-pro-3-sp/ (Accessed: 20 May 2024).
46. van den Oord, A. and Dieleman, S. (2016) *WaveNet: A generative model for raw audio* [Online]. Available at: https://deepmind.google/discover/blog/wavenet-a-generative-model-for-raw-audio/ (Accessed: 20 May 2024).
47. Wichern, G. (2017) *What the Machine Learning in RX 6 Advanced Means for the Future of Audio Repair Technology* [Online]. Available at: https://www.izotope.com/en/learn/what-the-machine-learning-in-rx-6-advanced-means-for-the-future-of-audio-repair-technology.html (Accessed: 20 May 2024).
48. Zhou, Y. et al (2018) *Visual to Sound: Generating Natural Sound for Videos in the Wild*, arXiv:1712.01393v2 [cs.CV] [Online]. Available at: https://arxiv.org/pdf/1712.01393/ (Accessed: 20 May 2024).
49. Zynaptiq (2013) *Zynaptiq Acquires Prosoniq Product Line And Technologies* [Online]. Available at: https://www.zynaptiq.com/more/zynaptiq-acquires-prosoniq-product-line-and-technologies/d106b1606afa940345ffb4a939c4588c/ (Accessed: 20 May 2024).

Filmography:

1. *People are not Goats* (1954) [DVD] Directed by Dragos Witkowski [Film]. București: Arhiva UNATC.

Dan-Ștefan Rucăreanu is a university lecturer at the I.L. Caragiale National University of Theatre and Film in Bucharest, in the Multimedia Department: Sound and Film Editing, teaching Sound Design for film, animation, new media, video games and film archive restoration. He is also the Study Programme Coordinator of Sound in the Multimedia Department. In parallel, he works as a sound designer and re-recording mixer in the film industry, participating in the making of a large number of films, documentaries and animations, many of which have been internationally awarded. He is a member of the European Film Academy and has been nationally awarded for sound design. With over 16 years of experience in the field of sound for cinematography and media, he is actively anchored in the domain of artificial intelligence for sound, each time looking for new methods involving the use of emerging AI technologies in his practice.

Laura Lăzărescu-Thois graduated in 2008 from the Multimedia, Sound and Film Editing Department of the Film Faculty of the I.L. Caragiale UNATC in Bucharest. In 2009 she became a student of the Doctoral School of UNATC, with a research topic relating to sound in the American animation film. She held two research grants in Berlin at the Hochschule für Film und Fernsehen "Konrad Wolf" in Potsdam-Babelsberg, Germany, and in June 2012 she obtained her PhD degree in the field of Cinematography and Media. In 2018 she finished a second study, BA in marketing at the Academy of Economic Studies in Bucharest. She works in both film and video production (as an editor, sound designer or director), and in the marketing field, being responsible for the strategy, the content or the online marketing for different festivals, agencies, or clients. Laura also published three books, scientific articles and market research articles concerning innovative educational methods and the modernization of the teaching process, which she presented at different international conferences. Currently she is an Associate Professor PhD, teaching at the Sound Department of the Film Faculty of UNATC.