Industry Corner
Debdeep Paul

In this "Industry Corner" column, we interview Dr. Nikolay Gaubitch, the Director of Research at Pindrop, where he leads the development of algorithms for future speech security technologies. He advises clients pre and post installation on the levels and types of phone fraud and he provides commentary on what is happening within the world of contact centre phone fraud. He received a Ph.D. in acoustic signal processing from Imperial College London in 2007. Between 2007 and 2012, he was a member of staff at Imperial College London where he managed the Centre for Law Enforcement Audio Research. Between 2012 and 2015, he was a Postdoctoral Researcher with the Signal and Information Processing Laboratory at Delft University of Technology where he worked on ad-hoc microphone arrays for speech enhancement in collaboration with Google.

In this interview, Dr. Gaubitch will give us a view of the threat landscape behind audio deepfakes and voice cloning technologies, how Pindrop is tackling these threats, what open research directions exist, and how academia-industry partnerships can help address these gaps. He concludes with advice for young researchers entering the field of deepfakes and deepfake detection on what topics/skills are needed. We hope you enjoy the interview!

---

1. Please give us a bit of background about yourself and how you ended up at Pindrop.

I obtained a PhD in acoustic signal processing with a thesis on speech dereverberation from Imperial College London in 2007. After that I worked as a post-doctoral researcher at Imperial College London focusing on forensic audio signal processing and at TU Delft on a Google-sponsored project on ad-hoc microphone arrays. One day in 2014, a recruiter from Pindrop reached out to me. I rarely respond to recruiters but something about the story of Pindrop and its mission resonated with me and before I knew it, I had entered the interview process. It was fascinating to see signal processing and machine learning in action for improving call centre security; remember that this is back in the day when your "identity" was verified over the phone with a few simple questions such as, "What is your mother's maiden name?". Another aspect that I found very exciting was that although Pindrop had existed in the US for a few years already and had some big banks as customers, there was no such technology available in Europe and we got to be the first to use technology to explore the telephony fraud landscape of European call centres.

2. Please give us a bit of background on Pindrop and how it has evolved over the last few years.

Pindrop started in 2011 with a focus on security and identity for every voice interaction, particularly to help enterprises answer the question "are we interacting with the right human?". This question has implications for organizations not only from a security perspective (stopping fraud attempts, avoiding negative publicity) but also from a customer experience standpoint (recognizing genuine customers and providing them a satisfactory experience). Pindrop created a multi-factor platform consisting of Protect, a fraud prevention solution and Passport, an authentication solution to help organizations answer the “right human” question. These platforms help the largest financial institutions, insurers and retailers in the world protect their contact centres from malicious actors and provide their genuine customers a superior experience through a passive and low friction authentication process. Since then, the Pindrop platforms have processed over 5 billion calls across all of our customers.

In 2023, as generative artificial intelligence (GenAI) became mainstream, it started posing threats in the form of synthetic audio or deepfakes. As deepfakes exploded both on the consumer side and in businesses, organizations began to struggle with a new question “are we interacting with a real human?”. At Pindrop Research, we had already worked on the topic of deepfakes and more generally on synthetic and modified speech detection for several years. And so when GenAI had become a mature technology, we were able to quickly address this new threat by creating a liveness detection solution called Pulse. Pulse is an AI driven audio deepfake detection solution that is modeled with over 20 million voice samples and with 120 text-to-speech engines and can identify synthetic audio with very high accuracy. It is capable of detecting not only presentation attacks but also voice cloning and real time voice conversion tactics and prevent companies from being scammed by malicious actors.

3. Deepfakes are starting to pose serious threats to companies, elections, information, etc. One example is the recent Joe Biden deepfake that Pindrop was able to not only detect, but also pinpoint which neural speech synthesizer generated it. Pindrop has recently released a report on the consumer sentiment around deepfakes and voice cloning. What were the major takeaways from this report?

Pindrop published “2023 Deepfake and Voice Clone Consumer Report”, in partnership with Voicebot.ai after surveying 2,000 U.S. consumers about their knowledge and feelings around deepfake and voice clone technology. Some of the key findings of the report include:

- People are learning about deepfake from social media
  Unsurprisingly, social media is leading the charge with channels like YouTube, TikTok, Instagram, and Facebook, followed by movies, documentaries, and the news media, where these technologies are used and covered when relevant headlines pop up. Awareness of deepfakes across these channels was slightly higher than voice clones.

- Consumers have a cautiously positive sentiment around deepfakes and voice clones
  Consumers find deepfakes to be funny and entertaining and are seen to improve realism and add creativity. But 90% of the population had some sort of concern around the technology.

- People hold slightly more pessimistic view of voice clones
  The report showed that many skewed either more positive or negative, like an inverted bell curve. The highest number for deepfakes was extremely positive and negative (22.3% each, respectively), with fewer people in the neutral middle (11.8%). Voice clones looked slightly

more pessimistic, with 21.6% in the extremely negative category and only 18.8% in the extremely positive category, perhaps because awareness is slightly lower.

- Consumer sentiment varies by industry
  Two-thirds of those surveyed knew about voice clones taking place in the banking industry. Voice clone awareness was also high in the politics and government and media industries. Consumers also feel that many industries such as banking, insurance, and healthcare are not doing enough to protect against the risks of deepfake technology.

- Awareness of deepfakes and voice clones varies by age and income
  Those aged over 60 showed a big drop in awareness. Concern around deepfakes and voice clones rises with income. The report also showed that consumers were very interested in voice authentication as a tool to protect against larger deepfake threats.

4. Within deepfakes and voice cloning detection, what are some open research problems that new students could focus on and try to help solve?

On the one hand, the field of synthetic speech detection is not a new field of research. On the other hand, there have been such major leaps in the performance of text-to-speech (TTS) and voice conversion (VC) technologies, that it may have rendered much of the past work irrelevant. As such there are many problems available from a research point of view and below is a list of some that I find particularly interesting:

- **Explainability in deepfake detection models**. This is an interesting and important question where a better understanding will be required to indicate why something is detected as deepfake. Another very important reason why this is a significant research problem is that there are and there will be an increasing number of examples of legitimate synthetic data. Thus, merely detecting a deepfake may not be enough in some applications – the distinction between legitimate and malicious synthetic data will have to be made.
- **Auxiliary technologies to protect against deepfakes.** Recent examples were the discussions around digital watermarking and to what extent these may be useful as protection against deepfakes. Future research could look at what other mechanisms may be developed.
- **The acoustic environment of audio recordings.** From my acoustic signal processing background, I find it particularly interesting to understand how acoustic features related to noise and reverberation and our current understanding of those may be used for more robust differentiation between real and synthetic data.

5. What role do you see academia playing in trying to help address the concerns around deepfakes and voice cloning? Is there room for academia-industry partnerships in this domain?

There is always room for academia-industry partnerships in any field and deepfakes and voice cloning are no exception. Here are three avenues that I believe are of importance. First, a notable example that has been popular in the past decade is that of organized challenges on specific topics. One such set of challenges that is very relevant to deepfake detection, as well as being a great example of academia-industry collaboration, is ASVspoof which has been running since 2015.

Second, there is an ongoing trend of large research investments in academia, supported by industry in the development of AI. One such recent example is the AI hub for generative models which is part of the £100m investment of UK Research & Innovation, spearheaded by UCL. This is a gathering of leading UK universities and industrial partners of which Pindrop is one.

Third, there are many recent developments in generative AI, among which TTS and VC, as well as deepfake detection which are relatively new to our society. Therefore there will be an increasing need for interdisciplinary collaboration between academia, industry and governments for the development of new standards, legislation and regulation around these new technologies. Examples of early such initiatives are the formation of the special interest group on security and privacy in speech communication as part of the International Speech Communication Association (ISCA) and even more so the recent development of the European Union AI Act.

6. What advice would you offer young researchers entering the field of deepfakes and deepfake detection, from the latest technical expertise that is/will be required, to the non-technical skills sought today by hiring managers at Pindrop and elsewhere?

Since we are on the topic of deepfakes and deepfake detection, my first piece of advice would be to think about the potential ethical and societal consequences of the technology we facilitate as a result of our research. We saw this with the latest rapid developments in TTS and VC systems that pushed the art of the possible to a new level. While these are exciting research and engineering achievements, they quickly also became a risk to social and political stability.

When it comes to the skills that one should think to acquire, I believe that one important aspect is to understand your domain well. For example, working with speech and audio, and the closely related topic of room acoustics, it is essential to gain some deeper understanding of the theory available. It is too easy sometimes to dive into the world of machine learning and rush to solve a particular problem by throwing plenty of data into a deep neural network and hope for the best. If you look under the hood of the successful methods, there is a combination of solid understanding of the domain and of machine learning. Moreover, a good understanding of your domain is essential for good communication with both other experts but more importantly with those less familiar with your work.

Learn to collaborate early - already during your doctoral studies. Being a good team player is important in many places in life but it is absolutely essential in industry.

Finally, work on a research topic that you are passionate about. Then, when you look for your next steps in your career, find something that captures your interest and your curiosity. Passion for the problem at hand will be the first solid impression that you will make on a hiring manager. Then, combined with a solid skill set both on a technical and collaborative level, you should have it all.