C5   AI-Based Audio Enhancement May Cause False Evidence

Eddy B. Brixen, BA*, EBB-Consult, Smorum DK-2765, DENMARK

Learning Overview: The goal of this presentation is to raise awareness to the new pitfalls in the art of voice enhancement. This presentation underlines the necessity of personnel's scientific education to avoid false evidence.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by demonstrating how Artificial Intelligence (AI) offers far more advanced possibilities regarding audio enhancement and how to avoid false conclusions.

Introduction: It is a common task to enhance low-quality speech recordings of various origins. Recordings from analog or digital wiretaps, surveillance systems, dispatch logging, dictaphones, memo recorders, etc. may end up at the audio forensics specialist's lab. Based on knowledge, training, and available equipment/software, the specialist may choose and apply different audio tools, primarily various filters, to enhance the recordings. Formerly, it was adequate to implement fixed filters or adaptive filters to reduce the recordings' background noise. Now AI offers far more advanced possibilities. Some algorithms are designed to take care of the noise in quite intelligent ways. However, other newer and more advanced algorithms look into signals recognized as speech. This is excellent—if speech is present!

Recently, a new type of recording has been introduced to audio forensics experts. These are recordings in which whispering or fragments of words seem to be recognized. At least, this is what the client states. Private clients and even clients from law enforcement may have the opportunity to acquire software that is extremely easy to use—or rather, from which it is easy to obtain a result. However, without the right training and the basic understanding of what the software actually does, these people may misinterpret the results.

AI-based dialog-isolating algorithms basically look for any voice content and emphasize this part of the signal. As an audio forensic expert, one should always detect whether or not recordings underwent any enhancement before arriving for analysis. Comments from the client like: "This speech was not audible until the filtering was performed" should always make one suspicious. The first time it is heard, there may be doubt as to whether or not speech is present, which one should not have.

The Experiment: After a number of challenging cases, this study carried out a series of experiments to identify the problem. At first, audio files were created by recording low-level speech close to the bottom of the dynamic range, applying linear recording as well as various versions of bit compression implemented in mobile devices. The speech was masked partly by quantization noise, partly by bitreduction artifacts, and partly by background noise (room noise). The files were enhanced applying Dialogue Isolate, a module of the widely used software iZotope RX8. The degree of processing was determined by the software's "Repair assistant," a tool suggesting the amount of treatment. The files were assessed by listeners, and by analyzing the Long-Time Average Spectra (LTAS), to find the degree of signal "voice-shaping."

The result of this test showed that syllables of spoken words—especially from cell phone recordings—may change to an unacceptable degree. Further, the process may generate unidentifiable speech-like sounds. It was not possible to identify a clear "voice shaping" profile of the signal.

After this part of the experiment, it became obvious to take it one step further. In this step, "voice enhancement" was applied to files not containing any speech. Different noises (pink noise, babble noise, low-level bit-reduction artifacts) were recorded using both linear quantization and bit compression. The recordings were assessed by listening. The result of this experiment exhibited a surprising effect: No recognizable words, of course. However, noise that fell within the frequency range of speech was emphasized, leaving speech-like sounds on the recording.

In cases where the question would be whether or not persons were present in a room, the described enhancement process could be hazardous to follow as it might lead to wrong conclusions.

Conclusion: This presentation describes new pitfalls in the art of voice enhancement. It underlines the necessity of personnel's scientific education to avoid false evidence. This presentation is supported by relevant sound samples from the experiment.

Audio Enhancement, AI (Artificial Intelligence), False Evidence

*Presenting Author