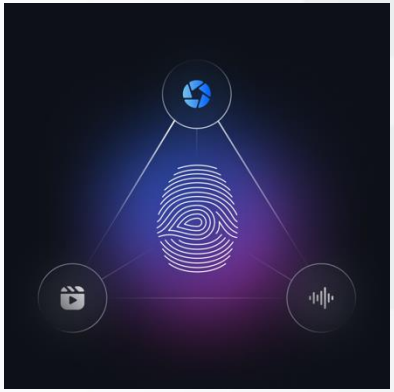
Audio Forensics

Areas of application, challenges and current developments

Gernot Schmied

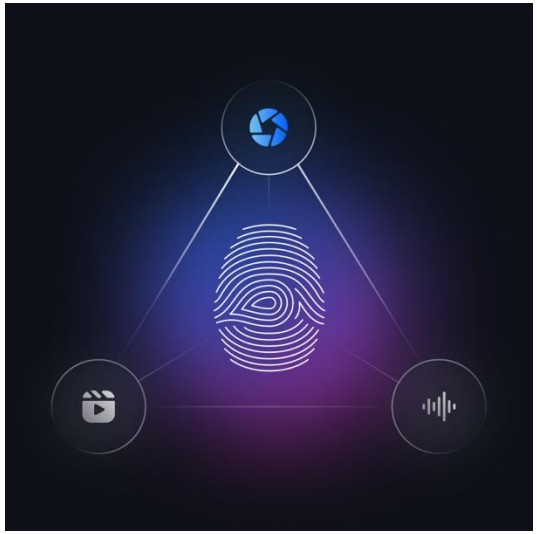
Multimedia Forensics Lab Vienna, AUSTRIA



tmt 33 TONMEISTERTAGUNG 2025

Gernot Schmied

- Living and working in Vienna, Austria
- Background in applied physics, computer science and an MBA
- Multimedia Forensics Lab Vienna – fascinated by the subject ☺
- Expert Witness in Court accredited by the Austrian Federal Ministry of Justice
- Particularly interested in legal aspects of digital forensics
- Practicing martial arts for quite some time

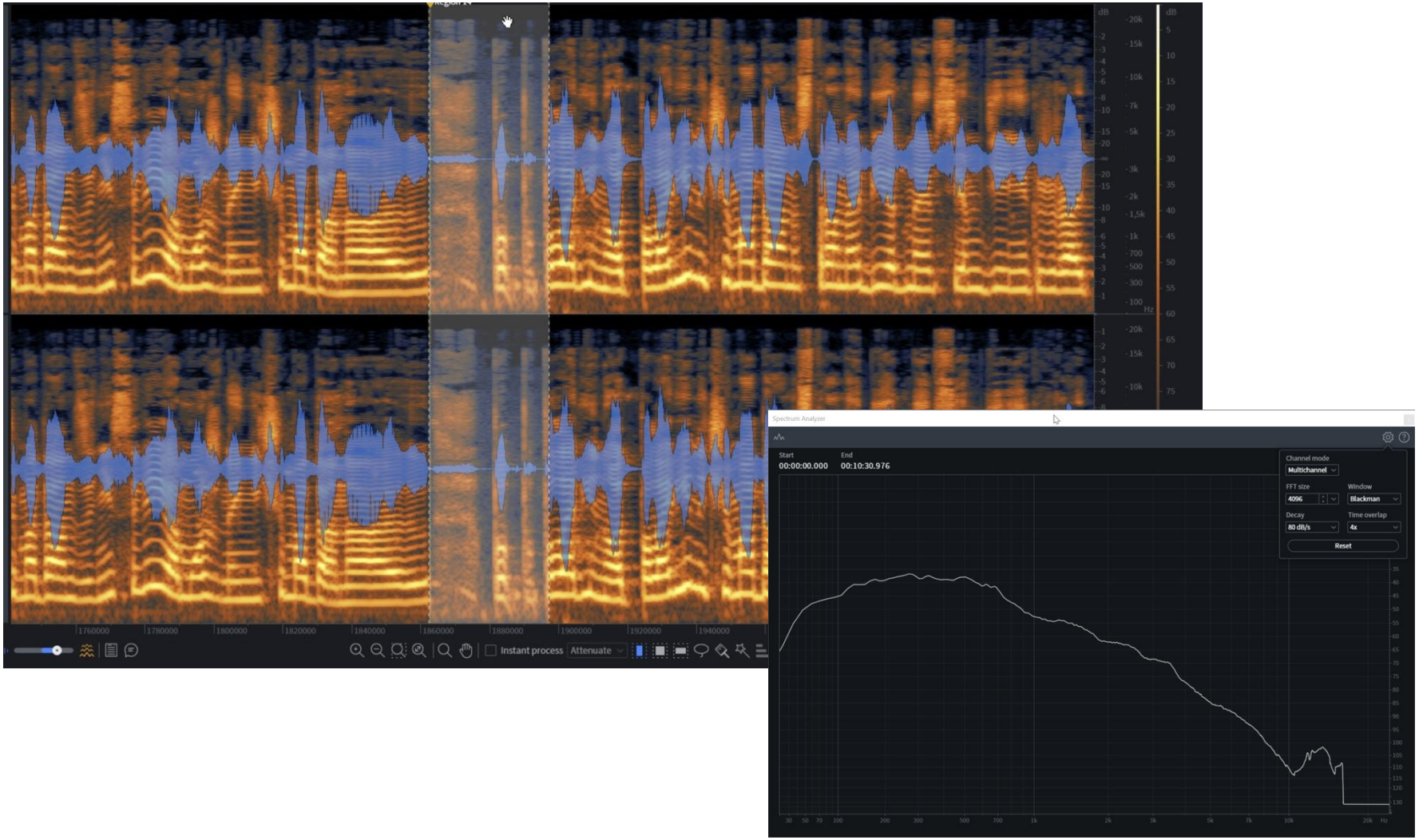


- **Audio Forensics**
  - fundamentally differs from Recording Arts, Audio Engineering, Mixing and Mastering. They do not have to worry about justifying every step of their work and process, if the results are pleasing and the **deliverables** live up to professional quality expectations.
  - is certainly not about **artistic expression** and aesthetics or optimizing for streaming, special replay devices or a concert hall audio design.
  - deals with **evidence items, preservation** and **casework** instead of projects, governed by **forensic standards** and documentation requirements. Our deliverables are **written expert witness reports** including **exhibits**, **exports, screenshots** and **attachments.**
  - often deals with exceptionally **poor recording quality** or lengthy recordings.
  - analysis and conclusions are **discussed** and **challenged** in legal proceedings.
  - occasionally does **peer reviews** with fellow labs and experts.
- However, the audio engineering body of knowledge, processes and toolbox are very valuable, audio forensics **uses an repurposes** a lot of the same tools, filters and plugins.
- Both disciplines overlap in the art of **audio restoration**, usually improving intelligibility of speech. Where we might differ, is especially in the use of **AI.**

Authenticating Evidence

- **Audio Restoration** is well understood and there are many tools available for that purpose, so is **transcription and translation** (speech to text).
- This is entirely different for **authentication** and **integrity verification** of audio evidence. For this very reason forensics adds programming, scripting and AI coding assistance (e.g. Matlab, Python, Claude, Gemini).
- We look for artifacts and clues to understand, whether a recording is an **original** or not (**provenance**), hence if it is **authentic** and the **integrity of the content** and its representation of real events unaltered and uncompromised.
- Authentication is more a discipline of science than technology, investigating **artifacts** and **traces** of editing and processing.
- In addition, **speaker** separation, identification, verification, **voice** comparison or speaker profiling is is more the domain of **linguists and phonetic specialists**.
- Obviously, it starts with **listening** to the evidence for a first impression and clues and an understanding of the **recording circumstances**, quality and length.
- Secondly, we look at metadata and **spectral representations**.
- With that knowledge check the **plausibility** of recording parameters and metadata.

Signal Analysis – Time and Frequency Domain
FFT Spectrogram and Spectrum



Forensic Paradigm and Expert Witness Approach

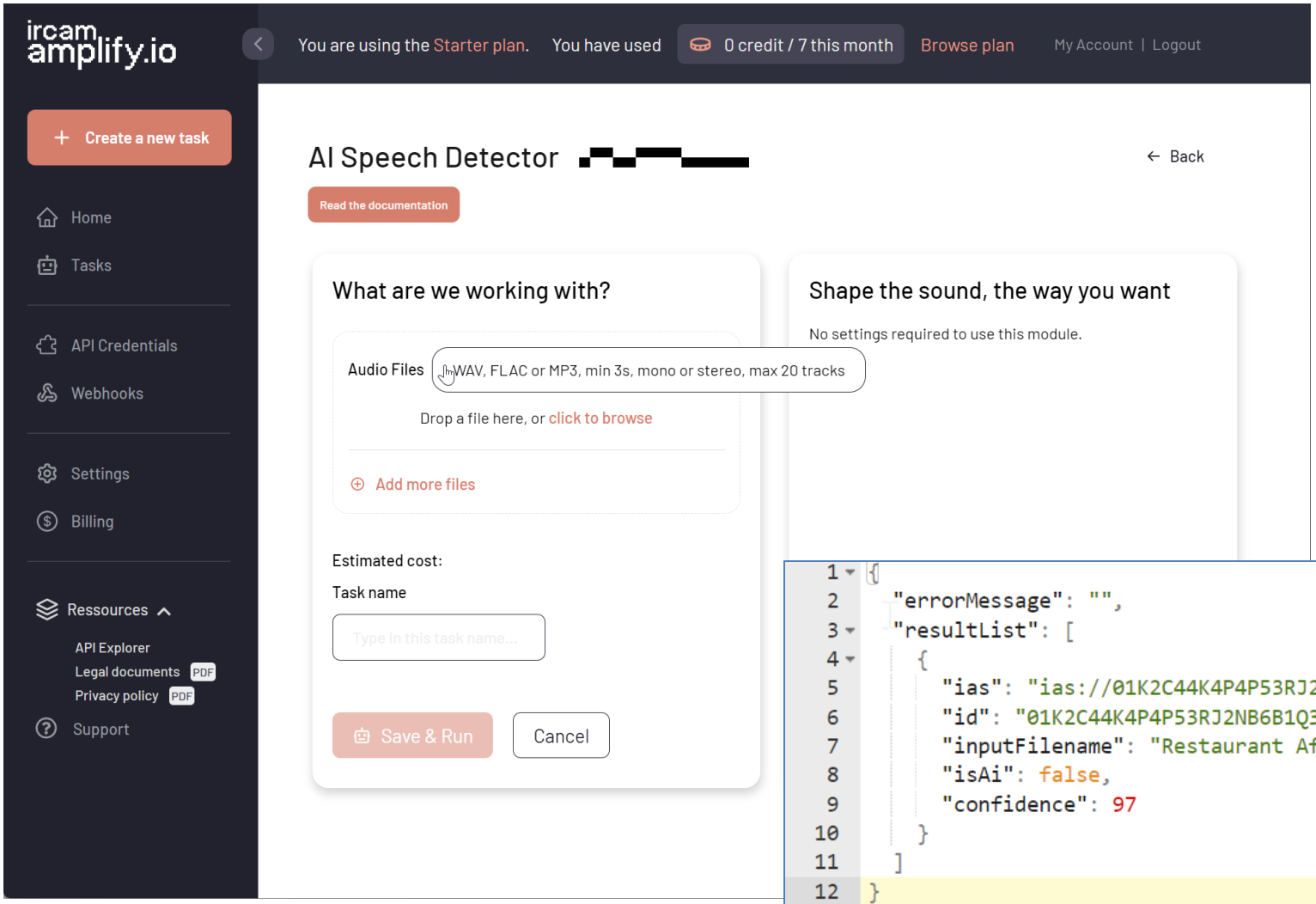
- It is of fundamental importance to preserve the **integrity of evidence** in any way possible, especially chain of custody and transfer.
- We want o be as close as possible to the **original recording** and are very much interested in identifying the **recording device**.
- **Reporting and expert witness written testimony:** The choices we make, and our decisions must be well documented, transparent, explainable, reproducible and repeatable. Document every step, method and parameterization of the analysis.
- **Manage expectations!** Most CSI movies have nothing to do with reality ☺
- **Reverse check:** All our conclusions can be traced back to findings, data, measurement and analysis, do not **speculate**! We either know with sufficient **confidence**, or we don't or the results are inconclusive. Never assume, always verify, do not fall victim to **bias**!
- **Never jump to quick or convenient conclusions!** Not every loss of integrity or failure to authenticate hast to be based on malicious intent, it also can happen accidentally and unintended.
- In general, evidence does not tell us about **motif or intent;** besides, we are not supposed to answer **legal questions**.

- This is closely related to the important question of conventional methods, filters and plugins versus the ones based on AI – *"domain specific specialized deep learning and neural networks".*
- So, **why bother** about AI explainability **when the results are impressive and convincing**? Because we do not know why, how and when it might fail, how well it was trained, accuracy and false positives/negatives, confidence and probability.
- AI models cannot easily be reduced to an **algorithm or mathematical formula**, hence much more difficult to explain and more elusive to the attempt.
- **Big problem – "black box":**
  Legal and forensic explainability to laymen about how it works without proper scientific foundation and whether it is reliable, accurate, trustworthy and the results reproducible and repeatable.
- **Even bigger problem:**
  The EU AI Act originates from product liability legislation. Even the conscious decision **to use AI** results in accountability, responsibility and liability and not just production, distribution or service providing.

Audio Deepfake Analysis

- **Audio deepfakes and voice cloning** has reached a level that we cannot trust or rely on our hearing senses anymore. Surprisingly, only short samples are required to produce convincing voice clones.
- Audio deepfake detection (online upload or via API key) rely on **AI methods to battle AI** generated deepfakes. This is not always smart and **orthogonal methods** important (structural analysis, CODECs, compression, Phase, DC-offset, Energy ..).
- **Establishing trust** in the verdict depends on transparency many services simply lack (*yes*, *no*):
  - Legal and technical explainability, AI act compliance, strong and detailed reporting
  - Training datasets, strength and weaknesses, meaningful benchmarking
  - Accuracy, false negatives/positives rate, confidence vs. probability, composite scores
  - Can deal with languages, dialects and idioms and simultaneous speakers?
  - Works with bad recording conditions and background noise?
- **Summary:** Audio deepfake detectors can at best give us a hint for further analysis, it is yet out of the question to give an expert witness statement solely based on that.

Audio Deepfake Analysis Example




```
{
  "errorMessage": "",
  "resultList": [
    {
      "ias": "ias://01K2C44K4P4P53RJ2NB6B1Q356",
      "id": "01K2C44K4P4P53RJ2NB6B1Q356",
      "inputFilename": "Restaurant After.mp3",
      "isAi": false,
      "confidence": 97
    }
  ]
}
```

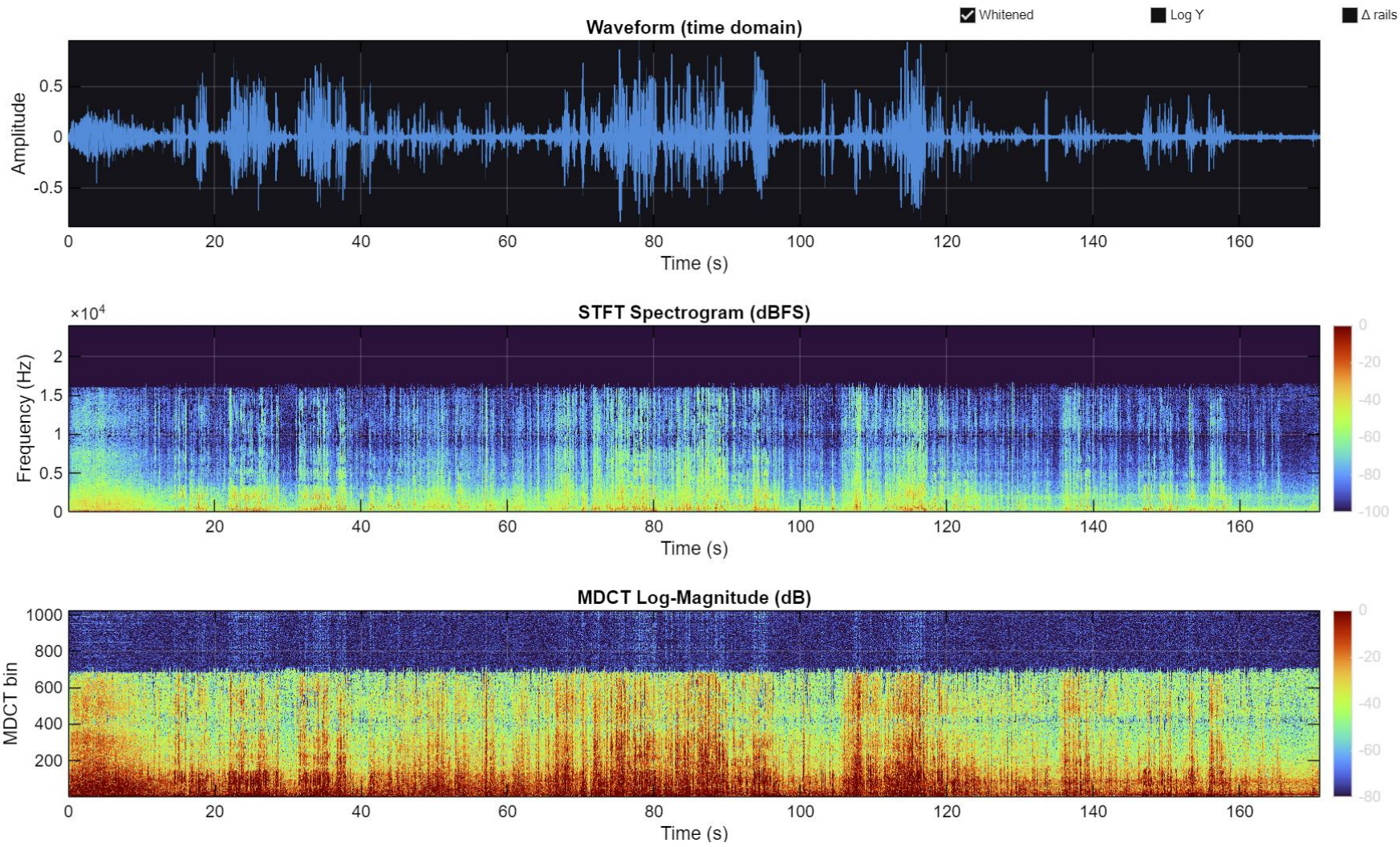
Choice of methods and tools

- The **choice of method and tools** depend on the recording circumstances and the quality of the evidence at hand (sample rate, SNR, length, language, varying noise, lossy/psycho-acoustic CODECs, clipping …) and consists of hardware and software.
  - DAWs, amps, headphones, filter, plugins and other applications
  - mix of commercial and Open Source, on-premises and secure Cloud Access
  - Matlab and Python scripting, Data science notebooks, Claude coding etc.
  - **FFMPEG** with frontends (ffMediaMaster WIN, ffWorks for MAC);
    FFMPEG8 integrates Whisper for on-premises local audio transcription.
- Some methods don't work well with bad quality evidence or simultaneous speakers. Some methods are calculation intensive and require splitting larger recordings into chunks.
- When dealing with forensic copies or deliverables, we always use **uncompressed and lossless .wav files!**
- It is possible to create **VST3 plugins** and GUIs from Matlab or Python code and integrate FFMPEG as well.
- Integration of online tools via **API Keys**, such as Deepfake detectors or stem separation, **batch processing** is also possible.

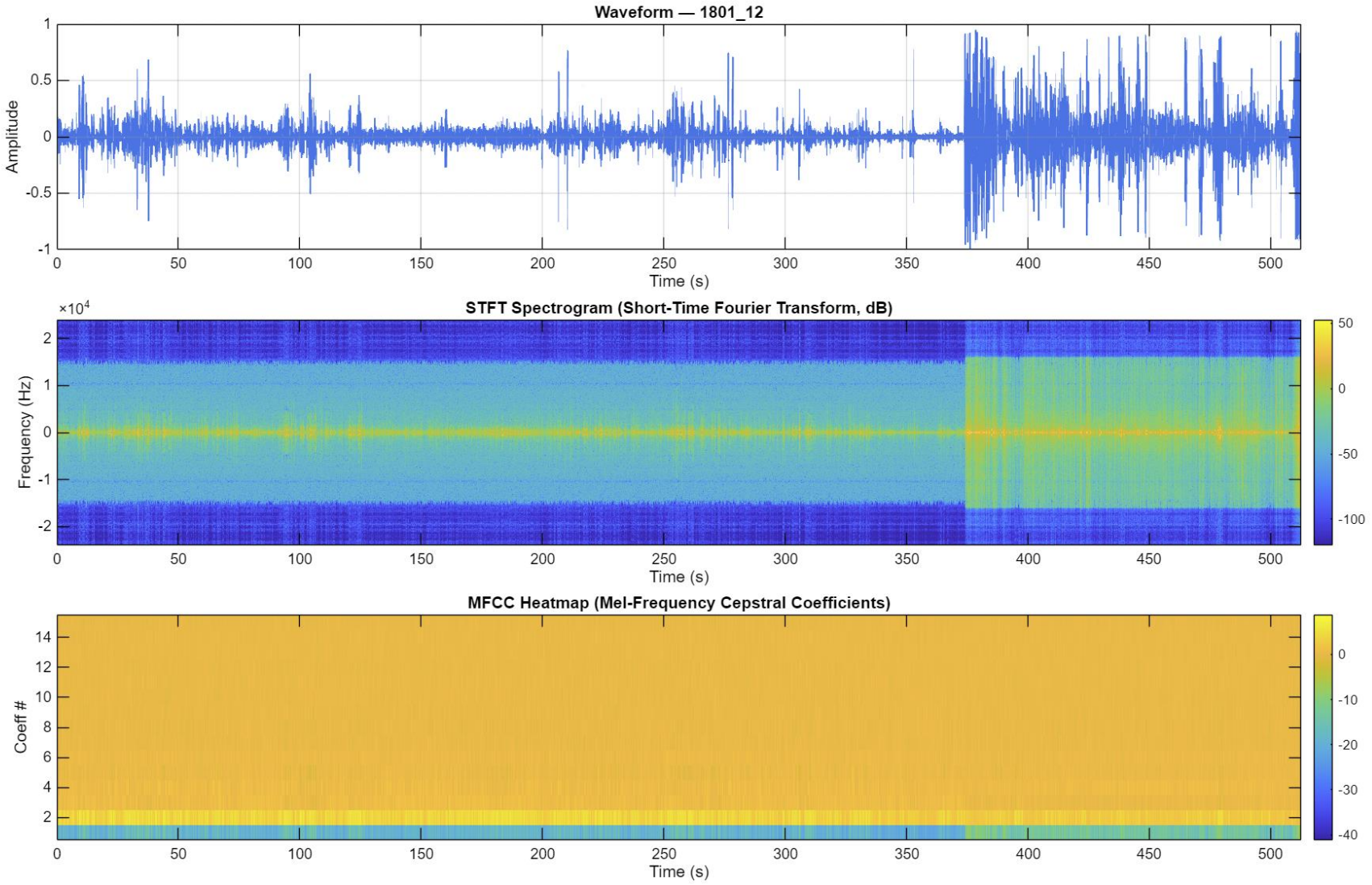
Advanced Methods of Analysis

- We are interested to detect **double compression/double encoding**, artifacts we would not expect to see in an original inconspicuous recording. Hence, we look at parameters and quantization that are related to encoding and compression, associated with reframing effects or altering the file structure.
- We are also interested to detect signs of **splicing or tampering or duplicate identical sequences** or in general any parameters that show **abrupt changes** we would as well not expect in an original recording (phase, dc-offset, energy …).
- Example approaches:
  - (M)DCT, (M)DWT (**wavelet** multi-resolution Analysis) and MFCC coefficient plots, histograms and heatmaps
  - Autocorrelation analysis, Frame offset analysis, Huffman coding analysis
  - ENF Analysis: 50/60Hz electromagnetic frequency inductive pickup
- **Challenges:** These methods are not robust under all conditions, qualitative interpretations from graphs and quantitative numerical analysis are difficult.
- **Trend:** Analysing the same parameters with AI methods such as (Convolutional) Neural networks or SVM (Support Vector Machines). AI can assist in and improve **interpretation** uncertainty.

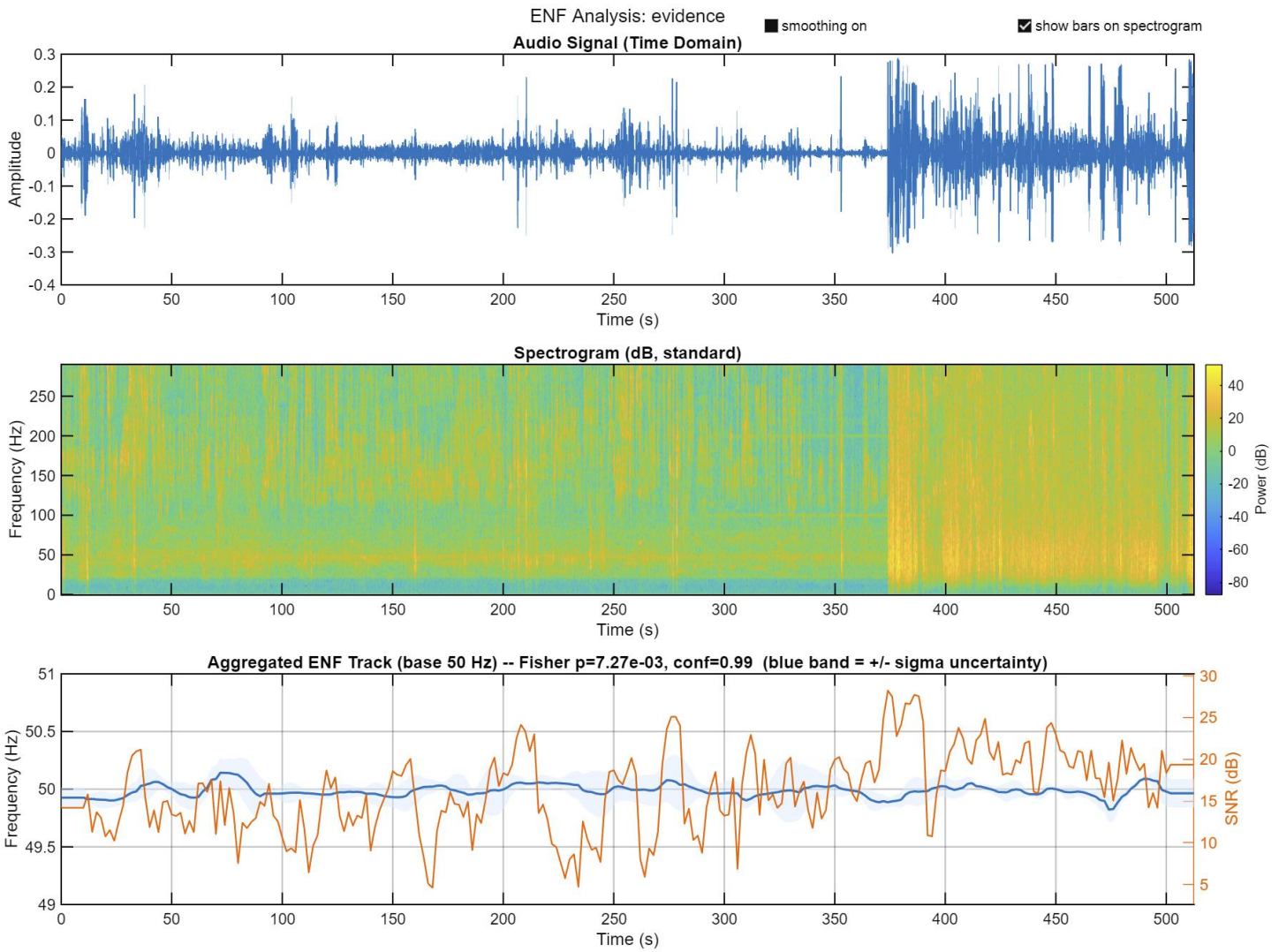
MDCT-Analysis (Modified Discrete Cosine Transformation)



MFCC-Analysis (Mel Frequency Cepstral Coefficients)



ENF-Analysis (Electrical Network Frequency pickup 50Hz)



Conclusions and wrap-up

- Regarding the quality and rapid improvement of deepfakes, **we cannot trust our hearing senses anymore**! This has fundamental consequences for introducing evidence in legal proceedings!
- If the analysis environment and the toolkit is unreliable, poorly understood or documented and non-deterministic, the result and expert witness testimony might be considered untrustworthy as well, worst case even inadmissible in legal proceedings. This is especially true for **AI!**
- In forensics it is mandatory to **document** due diligence efforts, workflow, tools, all the steps and filters and educated decisions made. Some tools assist that process with a detailed **history of interaction**. Expressing an **opinion** safely is also an art.
- AI can be useful and **acceptable** when consciously, transparently and competently assisting the forensic process and is deeply understood and documented.
- **Improvement, Restoration and Enhancement:**
  - Improvement OK, distortion of reality NOT OK
  - Minimally invasive and justifiable extent is OK
  - Alterations beyond necessity, aesthetic or artistical changes are NOT OK

gernot schmied
IT Sachverständiger & Ingenieurkonsulent

www.multimedia-forensik.com