# Scientific Working Group on Digital Evidence

## Best Practices for the Enhancement of Digital Audio

20-A-001-2.0

**The version of this document is in draft form and is being provided for comment by all interested parties for a minimum period of 60 days.**

### Disclaimer and Conditions Regarding Use of SWGDE Documents

SWGDE documents are developed by a consensus process that involves the best efforts of relevant subject matter experts, organizations, and input from other stakeholders to publish suggested best practices, practical guidance, technical positions, and educational information in the discipline of digital and multimedia forensics and related fields. No warranty or other representation as to SWGDE work product is made or intended.

As a condition to the use of this document (and the information contained herein) in any judicial, administrative, legislative, or other adjudicatory proceeding in the United States or elsewhere, the SWGDE requests notification by e-mail before or contemporaneous to the introduction of this document, or any portion thereof, as a marked exhibit offered for or moved into evidence in such proceeding. The notification should include: 1) The formal name of the proceeding, including docket number or similar identifier; 2) the name and location of the body conducting the hearing or proceeding; and 3) the name, mailing address (if available) and contact information of the party offering or moving the document into evidence. Subsequent to the use of this document in the proceeding please notify SWGDE as to the outcome of the matter. Notifications should be sent to secretary@swgde.org.

From time to time, SWGDE documents may be revised, updated, or sunsetted. Readers are advised to verify on the SWGDE website (https://www.swgde.org) they are utilizing the current version of this document. Prior versions of SWGDE documents are archived and available on the SWGDE website.

### Redistribution Policy

SWGDE grants permission for redistribution and use of all publicly posted documents created by SWGDE, provided that the following conditions are met:

1. Redistribution of documents or parts of documents must retain this SWGDE cover page containing the Disclaimer and Conditions of Use.
2. Neither the name of SWGDE nor the names of contributors may be used to endorse or promote products derived from its documents.
3. Any reference or quote from a SWGDE document must include the version number (or creation date) of the document and also indicate if the document is in a draft status.

### Requests for Modification

SWGDE encourages stakeholder participation in the preparation of documents. Suggestions for modifications are welcome and must be forwarded to the Secretary in writing at

secretary@swgde.org. The following information is required as a part of any suggested modification:

a) Submitter's name
b) Affiliation (agency/organization)
c) Address
d) Telephone number and email address
e) SWGDE Document title and version number
f) Change from (note document section number)
g) Change to (provide suggested text where appropriate; comments not including suggested text will not be considered)
h) Basis for suggested modification

**Intellectual Property**

Unauthorized use of the SWGDE logo or documents without written permission from SWGDE is a violation of our intellectual property rights.

Individuals may not misstate and/or over represent duties and responsibilities of SWGDE work. This includes claiming oneself as a contributing member without actively participating in SWGDE meetings; claiming oneself as an officer of SWGDE without serving as such; claiming sole authorship of a document; use the SWGDE logo on any material and/or curriculum vitae.

Any mention of specific products within SWGDE documents is for informational purposes only; it does not imply a recommendation or endorsement by SWGDE.

# Scientific Working Group on Digital Evidence

## Best Practices for the Enhancement of Digital Audio

## Table of Contents

# 1. Introduction

## 1.1 Purpose/Significance & Use

1.1.1 Audio enhancement is the processing and filtering of audio recordings to improve the signal quality and intelligibility of the signals of interest, such as speech, by attenuating noise or otherwise increasing the signal-to-noise ratio. [1]

1.1.2 Enhancement results must be repeatable and reproducible to the extent that they are perceptually consistent.

1.1.3 Enhancement may or may not improve the signals of interest, depending on the quality and technical characteristics of the submitted recording.

1.1.4 This document is intended to be a guide and summary of methods for forensic audio enhancement.

## 1.2 Scope

1.2.1 This document describes technical considerations and procedures to conduct forensic enhancement of digital audio.

1.2.2 This document contains recommendations for the review and analysis of an audio recording to assess the challenges to intelligibility and signal quality, the establishment of a processing workflow to address those challenges, and guidelines for applying various processing methods to improve intelligibility and signal quality.

1.2.3 This document does not cover every conceivable signal enhancement strategy.

1.2.4 This document assumes that the recording to be enhanced is a digital audio file handled according to *SWGDE Best Practices for Forensic Audio* [2].

1.2.5 This document does not address all safety concerns regarding the practice of audio enhancement. Repeated overexposure to noise at or above 85 dBA can cause permanent hearing loss, tinnitus, and difficulty understanding speech in noise [3]. Action should be taken to protect the hearing and well-being of examiners.

1.2.6 This document cannot replace knowledge, skills, or abilities acquired through education, training, and experience and is to be used in conjunction with professional judgment by individuals with such discipline-specific knowledge, skills, and abilities. Refer to *SWGDE Core Competencies for Forensic Audio* [4] for the minimum knowledge and abilities an examiner making audio enhancements should possess.

### 1.3 Limitations of Audio Enhancement

1.3.1 Forensic audio recordings are typically recorded in non-ideal situations and can suffer from strong distortions and a low signal-to-noise ratio. The goal of increasing intelligibility or signal quality may not be achievable.

1.3.2 Perceived signal quality and intelligibility can be evaluated differently by different listeners, as they are subjective functions of the auditory process.

1.3.3 Objective measures for evaluating intelligibility and signal quality based on perceptual models exist, but they may not be effective in forensic audio.

### 1.4 Summary of Practice

1.4.1 The overall audio enhancement process is comprised of the following stages:

1.4.1.1 Prepare a forensic working copy of the evidentiary material.

1.4.1.2 Locate the region of interest (ROI) that contains the target signal.

1.4.1.3 Assess the challenges to intelligibility and signal quality. Appendix A describes common challenges and proposed mitigation strategies.

1.4.1.4 Identify the processing necessary to mitigate the challenges identified.

1.4.1.5 Plan a workflow to apply the required processing without creating unwanted artifacts.

1.4.1.6 Apply the processing workflow, combined with necessary revisions to the workflow as they are identified.

1.4.1.7 Prepare the final results for distribution.

## 2. Pre-Examination

**2.1** Review the information provided by the submitter to determine the acoustic events recorded, region of interest, and target signal. If necessary, contact the submitter to clarify any issues or to obtain additional information. For example, the recording method and settings can be important for optimal audio enhancement. Document communications relevant to the request and examination.

**2.2** Parties submitting evidence to the laboratory might not be familiar with the best practices for audio enhancement. The laboratory should advise them of the best practices in this document and SWGDE Best Practices for Forensic Audio [2], to ensure that the most appropriate form of the evidence is submitted.

**2.3** The quality of the enhanced audio depends on the quality of the submitted recording. Processing applied to the audio prior to submitting it to the lab may limit the effectiveness of enhancement. If during the course of the examination it is found that an earlier generation recording may exist, contact the submitter.

**2.4** If the original audio recording is not provided, document your efforts to obtain the original evidence and peripheral items, as needed, and inform the submitter of any limitations imposed on the examination.

## 3. Assessment

**3.1** All examinations should be carried out on a working copy of the evidentiary materials or a version transcoded appropriately per the guidelines set forth in the "Preliminary Evidence Exam" section from SWGDE Best Practices for Forensic Audio [2].

**3.2** Conduct a technical evaluation using the following analyses to document the challenges affecting intelligibility or quality of the target signal. For more detailed information, see Appendix A – Challenges Table.

- Aural review, "critical listening"
- Waveform analysis for time-domain issues
- Multi-channel assessment, if applicable (e.g., amplitude differences, time offsets, content differences, phase differences)
- Spectrographic analysis of the frequency-domain with respect to time
- Spectral analysis of frequency content

**3.3** There is a risk of negatively affecting linguistically relevant information when processing speech in an unfamiliar language. Consider consulting with a fluent speaker.

**3.4** Document the regions of interest for enhancement.

**3.5** The challenges affecting intelligibility or signal quality can vary over the duration of the recording due to changes in recorder location, proximity of sound sources, or other factors. Document these changes as they may require different processing.

## 4. Workflow

**4.1** Determine strategies to mitigate the challenges identified during the assessment. Appendix A – Challenges Table describes mitigation strategies for common challenges.

4.1.1 Software applications may implement proprietary or unpublished algorithms. Their effectiveness should be tested before use.

4.1.2 Application of different processing methods to different time segments of recording may be necessary.

4.1.3 If the target signal is speech intended to be used in a speaker recognition process, only those mitigation strategies that do not impact speaker recognition results should be employed [5]. Refer to the OSAC Speaker Recognition Committee [6] for guidance.

**4.2** Decide an order to apply the selected mitigation strategies that maximizes the overall effectiveness of the enhancement.

4.2.1 Certain mitigation processes may be more effective when applied before or after others. The following order [7], while neither exhaustive nor compulsory, can be useful for most enhancement tasks [8] [9].

- Address distortions (e.g., de-click, de-clip)
- Source separation (e.g., spectral subtraction, reference channel cancellation)
- Attenuate continuous noises (e.g., noise reduction, equalization)
- Attenuate varying noises (e.g., adaptive noise reduction)
- De-reverberation
- Balance target signal characteristics (e.g., attenuation of sibilance and plosives)
- Gain correction (e.g., compensation for differences in talker loudness)

### 4.3 Apply the processing

4.3.1 Document the processes, settings, and applications used, with their versions, the time segments processed, and other relevant information in sufficient detail to repeat or reproduce the final results.

4.3.1.1 Application project files, history logs, and screen shots may serve as documentation of their content.

4.3.2 Establish optimal settings for each process.

4.3.2.1 Compare the process output to the input.

4.3.2.2 Assess progress towards increased intelligibility and listenability.

4.3.2.3 Avoid over-processing. Review filter residue (the signal components being removed) to avoid removing target signal components.

4.3.2.4 Compare different versions or iterations of processed audio.

4.3.3 If using multiple software applications, intermediate audio files should be maintained in an uncompressed format. Be aware of sample rate and bit depth quality and compatibility issues.

4.3.4 Listener (ear) fatigue occurs over extended periods of use as the physiological and psychoacoustic auditory system becomes overused and strained. Take breaks to avoid ear fatigue.

4.3.5 If, during the course of analysis, it is found that the original request is not achievable, communicate with the submitter to determine if the examination should continue.

### 4.4 Review

4.4.1 Perform a final review comparing the result to the unprocessed original.

4.4.2 The auditory system should be rested before review.

4.4.3 If the result is not satisfactory, revise the mitigation strategy and reprocess unless further processing will not improve the result.

### 4.5 Preparation of Results

4.5.1 Refer to the "Results of Examination" section of *SWGDE Best Practices for Forensic Audio* [2] for returning the results to the submitter.

4.5.2 Production of multiple enhanced results may be helpful to the submitter, e.g., cases in which overlapping signals of interest require different processing approaches.

4.5.3 It may be useful to accompany the output product with a copy of the submitted audio in a universally-playable format.

4.5.4 Advise the submitter regarding the use of proper audio playback equipment (e.g., over-ear headphones) when reviewing material individually as well as in the courtroom. Refer to *SWGDE Practical Considerations for Submission and Presentation of Multimedia Evidence in Court* [10].

## 5. Terminology

**5.1 intelligibility, n**, n—property of a signal representing its comprehensibility or its capability of being understood. With regard to speech, it is the proportion of a speaker's output that a listener can readily understand.

**5.2 listenability, n**—property of audio representing the ability to be listened to without discomfort.

**5.3 signal quality, n**—accuracy to which a signal represents acoustic events. Can be described by the ratio of signal to noise (SNR), or other measures.

## 6. Appendix A – Challenges Table

| Challenge | Explanation | Proposed Mitigation Strategies |
|---|---|---|
| Direct current (DC) offset | Mean bias (offset) in sample amplitude values away from zero. One or more components in the system add DC voltage(s) to a recorded signal resulting in a waveform that is not centered on the x-axis or 0 sample value. | DC offset removal filter; mean subtraction |
| Clipping | A distortion occurring when the signal exceeds a maximum threshold value or a recording system's capacity to represent it. | Sample interpolation (i.e., "de-clip") |
| Drop-out | Loss of acoustic information or missing samples, often represented as a constant low-level or zero-level quantization. | None, but document its location and duration |
| Impulses (pops, clicks, buzz) | Short duration amplitude spikes or a sequence of pulses. May be acoustic or electrical events. May be isolated, periodic, or sporadic. | Time-domain or frequency-domain interpolation ("de-click"); dynamic range compression; limiting; attenuation |

| Challenge | Explanation | Proposed Mitigation Strategies |
|---|---|---|
| Cell phone interference (GSM, CDMA) | A type of repetitive impulse noise caused by a digital transmitter coupling with a microphone or speaker, generating an interfering pulse train which can be consistent in shape and frequency. If powerful enough, it can cause distortion. | Time-domain or frequency-domain interpolation ("de-click", "de-buzz"); specialized tools that are designed to exploit the structure of the pulse train; matched filtering |
| Wind noise | Wind can generate low-frequency noise. If powerful enough, it can cause distortion. | Dedicated trained filter; dynamic range compression or limiting; gain correction; high pass filter |
| Rustle | Fabric rubbing on or near the microphone can introduce broadband noise and impulses. If powerful enough, it can cause distortion. | Dedicated trained filter; dynamic range compression or limiting; gain correction |
| Low level recording | Could be caused by the low-level input settings of the recording system, characteristics of the source, poor transmittal path, or equipment malfunctions | Amplitude gain; normalization; dynamics processing (e.g., compression, limiting) |
| Signals of interest (e.g., talkers) at different recording levels | For example, a telephone recording where one talker is consistently louder than the other. | Dynamics processing (e.g., compression, limiting, a filter designed for voice leveling) |
| Limited frequency response | Could be caused by an obscured microphone or poor quality recording system or transmittal path (e.g., muffled signal) | Equalization |
| Stationary tones | May be singular or harmonic (e.g., hum). Frequencies are generally stable. | Notch filter (stationary or adaptive); spectral subtraction, comb filter (hum); spectral editing; spectral inverse |
| Non-stationary tones | May be singular or harmonic, with frequencies varying over time. Examples include whistles, sirens, and instrumental music. | Adaptive notch filters; spectral editing; adaptive filters |
| Rumble | Low frequency noise, usually below 50 Hz, caused by environmental conditions | High pass filter; band pass filter; adaptive filters; equalization |
| Broadband noise | Usually has audible components at all audio frequencies or within a broad range of frequencies, such as air-conditioner noise, mechanical noise, street noise, noise floor of the recording system | Dedicated or adaptive broadband noise filter ("de-hiss"); dedicated or adaptive spectral inverse filter; band pass filter; speech, noise, or source separation which may utilize deep learning filters (trained neural networks) |
| Mixed sound sources | Mixture of audio sources including voices, music, and other sounds which have overlapping spectral components (e.g., cocktail party effect). | Source separation, with or without one or more reference channels, based on: (1) adaptive filters; (2) spectral subtraction; or (3) deep learning filters (trained neural networks) |
| Reverberation | Convolved signals that affect the target signal's intelligibility. | De-reverberation; adaptive filters |
| Speech rate | Individual's rate of speech causes unintelligible or disputed utterances. | Time-stretch with maintained pitch |
| Harsh high-frequency consonants | Consonants such as "s", "f" | Attenuation of sibilants (i.e., "de-ess"); band-limited dynamic compression |

| Challenge | Explanation | Proposed Mitigation Strategies |
|---|---|---|
| Plosive consonants | Consonants such as “b”, “p” | Attenuation of plosives; dynamic compression |

## 7. References

[1] ASTM International, *E2916 Standard Terminology for Digital and Multimedia Evidence Examination*, West Conshohocken, PA: ASTM, 2019.

[2] Scientific Working Group on Digital Evidence, *SWGDE Best Practices for Forensic Audio*, SWGDE, 2022.

[3] C. L. Themann and E. A. Masterson, "Occupational noise exposure: a review of its effects, epidemiology, and impact with recommendations for reducing its burde," *The Journal of the Acoustical Society of America*, vol. 146, no. 5, p. 3879, 2019.

[4] Scientific Working Group on Digital Evidence, *SWGDE Core Competencies for Forensic Audio*, SWGDE, 2023.

[5] H. J. Künzel and P. Alexander, "Forensic Automatic Speaker Recognition with Degraded and Enhanced Speech," *Journal of the Audio Engineering Society*, vol. 62, no. 4, pp. 244-253, April 2014.

[6] The Organization of Scientific Area Committees for Forensic Science, "Speaker Recognition Subcommittee," [Online]. Available: https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/speaker-recognition-subcommittee. [Accessed 15 01 2020].

[7] J. Zjalic, *A Proposed Framework for Forensic Audio Enhancement*, Denver, CO: University of Colorado Denver, 2017.

[8] B. E. Koenig, D. S. Lacey and S. A. Killion, "Forensic enhancement of digital audio recordings," *Journal of the Audio Engineering Society*, vol. 55, no. 5, pp. 352-371, 2007.

[9] C. Grigoras and J. Smith, Audio Enhancement and Authentication, in Encyclopedia of Forensic Sciences, Second Edition, Elsevier, 2013.

[10] Scientific Working Group on Digital Evidence, *Practical Considerations for Submission and Presentation of Multimedia Evidence in Court*, SWGDE, 2020.

## History

| Revision | Issue Date | History |
|---|---|---|
| 1.0 DRAFT | 1/16/2020 | Submitted for approval to release for public comment. |
| 1.1 DRAFT | 6/1/2020 | Submitted for approval to release for public comment. |
| 1.2 DRAFT | 9/16/2020 | Technical and editorial changes to Section 3. |
| 2.0 DRAFT | 6/15/2023 | Update document format and numbering. |
| 2.0 | 7/25/2023 | Formatted for posting after SWGDE membership voted to release as a Draft for Public Comment. |