



IEEE Standard Adoption of Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) Technical Specification Context-based Audio Enhanced (CAE) Version 2.1

Developed by the
Collaborative Activities Governance Board

IEEE SA Industry Affiliated Network (IAN)
Base specification contributed by MPAI community



IEEE Std 3302™-2024

IEEE Standard Adoption of Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) Technical Specification Context-based Audio Enhancement (CAE) Version 2.1

Developed by the

Collaborative Activities Governance Board
of the
IEEE SA Board of Governors

Approved 11 December 2024

IEEE SA Standards Board

Abstract: This standard adopts MPAA Technical Specification Version 2.1 as an IEEE Standard. The Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAA) Technical Specification Context-based Audio Enhancement (CAE) Version 2.1 collects use cases that improve the user experience of audio applications using technologies that act on the input audio content using context information. The coverage of MPAA-CAE use cases includes entertainment, communication, teleconferencing, gaming, post-production, restoration, etc., in a variety of contexts such as in the home, in the car, on the go, and in the studio. The technologies are specified as AI Workflows, AI Modules, and Data Types. The currently specified use cases are Emotion Enhanced Speech (EES), Audio Recording Preservation (ARP), Speech Restoration System (SSR), and Enhanced Audioconference Experience (EAE).

Keywords: adoption, AI module, AI workflow, ARP, audio recording preservation, EAE, EES, emotion enhanced speech, enhanced audio-conference experience, IEEE 3302™, MPAA-CAE

The Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, New York, NY 10016-5997, USA

Copyright © 2025 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved. Published 6 June 2025. Printed in the United States of America.

IEEE is a registered trademark in the U.S. Patent & Trademark Office, owned by The Institute of Electrical and Electronics Engineers, Incorporated.

PDF: ISBN 979-8-8557-2180-5 STD27918
Print: ISBN 979-8-8557-2181-2 STDPD27918

IEEE prohibits discrimination, harassment, and bullying.

For more information, visit <https://www.ieee.org/about/corporate/governance/p9-26.html>.

No part of this publication may be reproduced in any form, in an electronic retrieval system or otherwise, without the prior written permission of the publisher.

Important Notices and Disclaimers Concerning IEEE Standards Documents

IEEE Standards documents are made available for use subject to important notices and legal disclaimers. These notices and disclaimers, or a reference to this page (<https://standards.ieee.org/ipr/disclaimers.html>), appear in all IEEE standards and may be found under the heading “Important Notices and Disclaimers Concerning IEEE Standards Documents.”

Notice and Disclaimer of Liability Concerning the Use of IEEE Standards Documents

IEEE Standards documents are developed within IEEE Societies and subcommittees of IEEE Standards Association (IEEE SA) Board of Governors. IEEE develops its standards through an accredited consensus development process, which brings together volunteers representing varied viewpoints and interests to achieve the final product. IEEE standards are documents developed by volunteers with scientific, academic, and industry-based expertise in technical working groups. Volunteers involved in technical working groups are not necessarily members of IEEE or IEEE SA and participate without compensation from IEEE. While IEEE administers the process and establishes rules to promote fairness in the consensus development process, IEEE does not independently evaluate, test, or verify the accuracy of any of the information or the soundness of any judgments contained in its standards.

IEEE makes no warranties or representations concerning its standards, and expressly disclaims all warranties, express or implied, concerning all standards, including but not limited to the warranties of merchantability, fitness for a particular purpose and non-infringement. IEEE Standards documents do not guarantee safety, security, health, or environmental protection, or compliance with law, or guarantee against interference with or from other devices or networks. In addition, IEEE does not warrant or represent that the use of the material contained in its standards is free from patent infringement. IEEE Standards documents are supplied “AS IS” and “WITH ALL FAULTS.”

Use of an IEEE standard is wholly voluntary. The existence of an IEEE standard does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to the scope of the IEEE standard. Furthermore, the viewpoint expressed at the time a standard is approved and issued is subject to change brought about through developments in the state of the art and comments received from users of the standard.

In publishing and making its standards available, IEEE is not suggesting or rendering professional or other services for, or on behalf of, any person or entity, nor is IEEE undertaking to perform any duty owed by any other person or entity to another. Any person utilizing any IEEE Standards document should rely upon their own independent judgment in the exercise of reasonable care in any given circumstances or, as appropriate, seek the advice of a competent professional in determining the appropriateness of a given IEEE standard.

IN NO EVENT SHALL IEEE BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: THE NEED TO PROCURE SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Translations

The IEEE consensus balloting process involves the review of documents in English only. In the event that an IEEE standard is translated, only the English language version published by IEEE is the approved IEEE standard.

Use by artificial intelligence systems

In no event shall material in any IEEE Standards documents be used for the purpose of creating, training, enhancing, developing, maintaining, or contributing to any artificial intelligence systems without the express, written consent of IEEE SA in advance. “Artificial intelligence” refers to any software, application, or other system that uses artificial intelligence, machine learning, or similar technologies, to analyze, train, process, or generate content. Requests for consent can be submitted using the Contact Us form.

Official statements

A statement, written or oral, that is not processed in accordance with the IEEE SA Standards Board Operations Manual is not, and shall not be considered or inferred to be, the official position of IEEE or any of its committees and shall not be considered to be, or be relied upon as, a formal position of IEEE or IEEE SA. At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that the presenter’s views should be considered the personal views of that individual rather than the formal position of IEEE, IEEE SA, the Standards Committee, or the Working Group. Statements made by volunteers may not represent the formal position of their employer(s) or affiliation(s). News releases about IEEE standards issued by entities other than IEEE SA should be considered the view of the entity issuing the release rather than the formal position of IEEE or IEEE SA.

Comments on standards

Comments for revision of IEEE Standards documents are welcome from any interested party, regardless of membership affiliation with IEEE or IEEE SA. However, **IEEE does not provide interpretations, consulting information, or advice pertaining to IEEE Standards documents.**

Suggestions for changes in documents should be in the form of a proposed change of text, together with appropriate supporting comments. Since IEEE standards represent a consensus of concerned interests, it is important that any responses to comments and questions also receive the concurrence of a balance of interests. For this reason, IEEE and the members of its Societies and subcommittees of the IEEE SA Board of Governors are not able to provide an instant response to comments or questions, except in those cases where the matter has previously been addressed. For the same reason, IEEE does not respond to interpretation requests. Any person who would like to participate in evaluating comments or revisions to an IEEE standard is welcome to join the relevant IEEE SA working group. You can indicate interest in a working group using the Interests tab in the Manage Profile & Interests area of the [IEEE SA myProject system](#).¹ An IEEE Account is needed to access the application.

Comments on standards should be submitted using the [Contact Us](#) form.²

¹ Available at: <https://development.standards.ieee.org/myproject-web/public/view.html#landing>.

² Available at: <https://standards.ieee.org/about/contact/>.

Laws and regulations

Users of IEEE Standards documents should consult all applicable laws and regulations. Compliance with the provisions of any IEEE Standards document does not constitute compliance to any applicable regulatory requirements. Implementers of the standard are responsible for observing or referring to the applicable regulatory requirements. IEEE does not, by the publication of its standards, intend to urge action that is not in compliance with applicable laws, and these documents may not be construed as doing so.

Data privacy

Users of IEEE Standards documents should evaluate the standards for considerations of data privacy and data ownership in the context of assessing and using the standards in compliance with applicable laws and regulations.

Copyrights

IEEE draft and approved standards are copyrighted by IEEE under U.S. and international copyright laws. They are made available by IEEE and are adopted for a wide variety of both public and private uses. These include both use by reference, in laws and regulations, and use in private self-regulation, standardization, and the promotion of engineering practices and methods. By making these documents available for use and adoption by public authorities and private users, neither IEEE nor its licensors waive any rights in copyright to the documents.

Photocopies

Subject to payment of the appropriate licensing fees, IEEE will grant users a limited, non-exclusive license to photocopy portions of any individual standard for company or organizational internal use or individual, non-commercial use only. To arrange for payment of licensing fees, please contact Copyright Clearance Center, Customer Service, 222 Rosewood Drive, Danvers, MA 01923 USA; +1 978 750 8400; <https://www.copyright.com/>. Permission to photocopy portions of any individual standard for educational classroom use can also be obtained through the Copyright Clearance Center.

Updating of IEEE Standards documents

Users of IEEE Standards documents should be aware that these documents may be superseded at any time by the issuance of new editions or may be amended from time to time through the issuance of amendments, corrigenda, or errata. An official IEEE document at any point in time consists of the current edition of the document together with any amendments, corrigenda, or errata then in effect.

Every IEEE standard is subjected to review at least every 10 years. When a document is more than 10 years old and has not undergone a revision process, it is reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art. Users are cautioned to check to determine that they have the latest edition of any IEEE standard.

In order to determine whether a given document is the current edition and whether it has been amended through the issuance of amendments, corrigenda, or errata, visit [IEEE Xplore](#) or [contact IEEE](#).³ For more information about the IEEE SA or IEEE's standards development process, visit the IEEE SA Website.

Errata

Errata, if any, for all IEEE standards can be accessed on the [IEEE SA Website](#).⁴ Search for standard number and year of approval to access the web page of the published standard. Errata links are located under the Additional Resources Details section. Errata are also available in [IEEE Xplore](#). Users are encouraged to periodically check for errata.

Patents

IEEE standards are developed in compliance with the [IEEE SA Patent Policy](#).⁵

Attention is called to the possibility that implementation of this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. If a patent holder or patent applicant has filed a statement of assurance via an Accepted Letter of Assurance, then the statement is listed on the IEEE SA Website at <https://standards.ieee.org/about/sasb/patcom/patents.html>. Letters of Assurance may indicate whether the Submitter is willing or unwilling to grant licenses under patent rights without compensation or under reasonable rates, with reasonable terms and conditions that are demonstrably free of any unfair discrimination to applicants desiring to obtain such licenses.

Essential Patent Claims may exist for which a Letter of Assurance has not been received. The IEEE is not responsible for identifying Essential Patent Claims for which a license may be required, for conducting inquiries into the legal validity or scope of Patents Claims, or determining whether any licensing terms or conditions provided in connection with submission of a Letter of Assurance, if any, or in any licensing agreements are reasonable or non-discriminatory. Users of this standard are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. Further information may be obtained from the IEEE Standards Association.

IMPORTANT NOTICE

Technologies, application of technologies, and recommended procedures in various industries evolve over time. The IEEE standards development process allows participants to review developments in industries, technologies, and practices, and to determine what, if any, updates should be made to the IEEE standard. During this evolution, the technologies and recommendations in IEEE standards may be implemented in ways not foreseen during the standard's development. IEEE standards development activities consider research and information presented to the standards development group in developing any safety recommendations. Other information about safety practices, changes in technology or technology implementation, or impact by peripheral systems also may be pertinent to safety considerations during implementation of the standard. Implementers and users of IEEE Standards documents are responsible for determining and complying with all appropriate safety, security, environmental, health, data privacy, and interference protection practices and all applicable laws and regulations.

³ Available at: <https://ieeexplore.ieee.org/browse/standards/collection/ieee>.

⁴ Available at: <https://standards.ieee.org/standard/index.html>.

⁵ Available at: <https://standards.ieee.org/about/sasb/patcom/materials.html>.

Participants

At the time this standard was completed, the Artificial Intelligence Framework Working Group had the following membership:

Marina Bosi, *Chair*
Stephen Dukes, *Vice Chair*
Yu Yuan, *Secretary*

<i>Organization Represented</i>	<i>Name of Representative</i>
0xSenses Corporation	Yu Yuan
1stCycle Corporation	Johnny Lin
Beijing Tsinghua Changung Hospital	Bin Yang
Imaginary Universes, LLC.....	Stephen Dukes
Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI)	Marina Bosi
PetroChina Planning and Engineering Institute	Pan Kai

The following members of the entity Standards Association balloting group voted on this standard. Balloters may have voted for approval, disapproval, or abstention.

0xSenses Corporation	Moving Picture, Audio and Data	Tongyuan Software Co., Ltd.
Imaginary Universes, LLC	Coding by Artificial	
Intel	Intelligence (MPAI)	

When the IEEE SA Standards Board approved this standard on 11 December 2024, it had the following membership:

David J. Law, *Chair*
Jon Walter Rosdahl, *Vice Chair*
Gary Hoffman, *Past Chair*
Alpesh Shah, *Secretary*

Sara R. Biyabani	Hao Hu	Paul Nikolich
Ted Burse	Yousef Kimiagar	Robby Robson
Stephen Dukes	Joseph L. Koepfinger*	Lei Wang
Doug Edwards	Howard Li	F. Keith Waters
J. Travis Griffith	Xiaohui Liu	Sha Wei
Guido R. Hiertz	John Haiying Lu	Philip B. Winston
Ronald W. Hotchkiss	Kevin W. Lu	Don Wright
	Hiroshi Mano	

*Member Emeritus



Moving Picture, Audio and Data Coding
by Artificial Intelligence
www.mpai.community

MPAI Technical Specification

Context-based Audio Enhancement MPAI-CAE

V2.1

WARNING

Use of the technologies described in this Technical Specification may infringe patents, copyrights or intellectual property rights of MPAI Members or non-members.

MPAI and its Members accept no responsibility whatsoever for damages or liability, direct or consequential, which may result from use of this Technical Specification.

Readers are invited to review Annex 2 - Notices and Disclaimers.

Context-based Audio Enhancement (MPAI-CAE) V2.1

Contents

1	Introduction	4
2	Scope	6
3	Terms and Definitions	7
4	References	9
4.1	Normative References	9
4.2	Informative References	10
5	Use Cases	10
5.1	Emotion-Enhanced Speech (CAE-EES)	10
5.1.1	Functions	10
5.1.2	Reference Model	11
5.1.3	I/O data of AI Workflow	11
5.1.4	Functions of AI Modules.....	12
5.1.5	I/O Data of AI Modules	12
5.1.6	AIW, AIMs, and JSON Metadata	12
5.2	Audio Recording Preservation (CAE-ARP)	13
5.2.1	Functions	13
5.2.2	Reference Model	13
5.2.3	I/O data of AI Workflow	15
5.2.4	Functions of AI Modules.....	15
5.2.5	I/O Data of AI Modules	16
5.2.6	AIW, AIMs, and JSON Metadata	16
5.3	Speech Restoration System (CAE-SRS)	17
5.3.1	Functions	17
5.3.2	Reference Model	17
5.3.3	I/O Data of AI Workflow	18
5.3.4	Functions of AI Modules.....	19
5.3.5	I/O Data of AI Modules	19
5.3.6	AIW, AIMs and JSON Metadata	19
5.4	Enhanced Audioconference Experience (CAE-EAE).....	20
5.4.1	Functions	20
5.4.2	Reference Model	20
5.4.3	I/O data of AI Workflow	21
5.4.4	Functions of AI Modules.....	21
5.4.5	I/O Data of AI Modules	23
5.4.6	AIW, AIMs, and JSON Metadata	23
5.5	Human-Connected Autonomous Vehicle (CAV) Interaction	24
5.5.1	Functions of Use Case.....	24
5.5.2	Reference Model	25
5.5.3	I/O Data of HCI AI Workflow	27
5.5.4	AIW, AIMs, and JSON Metadata	27
6	Audio Scene Description Composite AIM.....	28

6.1	Functions	28
6.2	Reference Model	28
6.3	I/O Data of Composite AIM.....	29
6.4	Functions of AIMs	29
6.5	I/O Data of AIMs	29
6.6	AIMs and JSON Metadata	30
7	Data Types.....	30
7.1	Access Copy Files	31
7.2	Audio Block	31
7.3	Audio File.....	31
7.4	Audio Object	31
7.4.1	Definition	31
7.4.2	Syntax.....	31
7.4.3	Semantics	32
7.5	Audio Scene Descriptors	33
7.5.1	Definition	33
7.5.2	Syntax.....	33
7.5.3	Semantics	34
7.6	Audio Scene Geometry	34
7.6.1	Definition	34
7.6.2	Syntax.....	35
7.6.3	Semantics	35
7.7	Audio Segment.....	36
7.8	Damaged List	36
7.8.1	Definition	36
7.8.2	Syntax.....	36
7.8.3	Semantics	37
7.9	Editing List.....	37
7.9.1	Definition	37
7.9.2	Syntax.....	37
7.9.3	Semantics	38
7.10	Emotion	39
7.11	Emotionless Speech.....	39
7.12	Enhanced Audio	39
7.13	Enhanced Transform Audio	40
7.14	Irregularity File	40
7.14.1	Definition	40
7.14.2	Syntax.....	40
7.14.3	Semantics	42
7.15	Irregularity Image.....	45
7.16	Microphone Array Audio	45
7.17	Microphone Array Geometry	45
7.17.1	Definition	45
7.17.2	Syntax.....	45
7.17.3	Semantics	47
7.18	Mode Selection.....	48
7.19	Multichannel Audio.....	48
7.20	Multichannel Audio-Stream	49

7.20.1 Definition	49
7.20.2 Syntax.....	49
7.20.3 Semantics	50
7.21 Neural Network Speech Model	50
7.22 Preservation Audio File.....	51
7.23 Preservation Audio-Visual File.....	51
7.24 Preservation Master Files	51
7.25 Speech Descriptors	51
7.25.1 Definition	51
7.25.2 Syntax.....	51
7.25.3 Semantics	52
7.26 Spherical Harmonic Decomposition	53
7.27 Transform Audio	53
7.28 Transform Multichannel Audio.....	53
7.29 Video	54
Annex 1 - MPAAI-wide terms and definitions	55
Annex 2 - Notices and Disclaimers Concerning MPAAI Standards (Informative).....	58
Annex 3 - - Examples (Informative)	60
3.1 Audio Scene Geometry	60
3.2 Damaged List	60
3.3 Editing List.....	60
3.4 Irregularity File	61
3.5 Microphone Array Geometry	62
3.6 Prosodic Speech Features.....	63
3.7 Neural Speech Features	64
Annex 4 - Communication Among AIM Implementors (Informative).....	65

1 Introduction

In recent years, Artificial Intelligence (AI) and related technologies, applied to a broad range of applications, have started affecting the life of millions of people and they are expected to do so even more in the future. As digital media standards have positively influenced industry and billions of people, so AI-based data coding standards are expected to have a similar positive impact. Indeed, research has shown that data coding with AI-based technologies is generally *more efficient* than with existing technologies for, e.g., compression and feature-based description.

However, some AI technologies may carry inherent risks, e.g., in terms of bias toward some classes of users. Therefore, the need for standardisation is more important and urgent than ever.

The international, unaffiliated, not-for-profit MPAAI – Moving Picture, Audio and Data Coding by Artificial Intelligence Standards Developing Organisation has the mission to develop *AI-enabled data coding standards*. MPAAI Application Standards enable the development of AI-based products, applications, and services.

As a part of its mission, MPAAI has developed standards operating procedures to enable a user of MPAAI implementations to make informed decision about their applicability. Central to this is the notion of Performance, defined as a set of attributes characterising a reliable and trustworthy implementation.

In the following, Terms beginning with a capital letter are defined in *Table 1* if they are specific to this Standard and in *Table 33* if they are common to all MPAI Standards.

Figure 1 depicts the MPAI-AIF Reference Model under which Implementations of MPAI Application Standards and user-defined MPAI-AIF conforming applications operate.

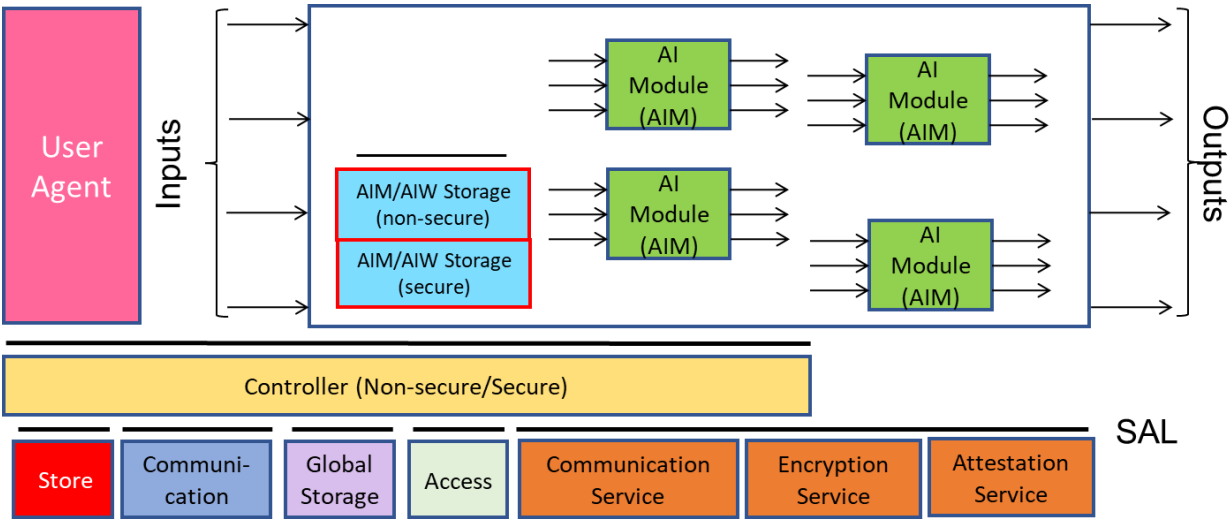


Figure 1 - The AI Framework (AIF) Reference Model and its Components

An AIF Implementation allows execution of AI Workflows (AIW), composed by basic processing elements called AI Modules (AIM).

MPAI Application Standards normatively specify Semantics and Syntax of the input and output data and the Function of the AIW and the AIMs, and the Connections between and among the AIMs of an AIW.

In particular, an AIM is defined by its Function and Data, but not by its internal architecture, which may be based on AI or data processing, and implemented in software, hardware or hybrid software and hardware technologies.

MPAI defines Interoperability as the ability to replace an AIW or an AIM Implementation with a functionally equivalent Implementation. MPAI also defines 3 Interoperability Levels of an AIW that executes an AIW. The AIW may have 3 Levels:

Level 1 – Implementer-specific and satisfying the MPAI-AIF Standard.

Level 2 – Specified by an MPAI Application Standard.

Level 3 – Specified by an MPAI Application Standard and certified by a Performance Assessor.

MPAI offers Users access to the promised benefits of AI with a guarantee of increased transparency, trust and reliability as the Interoperability Level of an Implementation moves from 1 to 3. Additional information on Interoperability Levels is provided in Annex 3.

The Chapters and Annexes of this Technical Specification are Normative unless they are labelled as Informative.

2 Scope

Technical Specification: Context-based Audio Enhancement (MPAI-CAE) V2.1 specifies technologies that improve the user experience for audio-related applications including entertainment, communication, teleconferencing, gaming, post-production, restoration etc. in a variety of contexts such as in the home, in the car, on-the-go, in the studio etc. using context information to act on the input audio content, and potentially deliver the processed output via an appropriate protocol. MPAI-CAE specifies four Use Cases and one Composite AIM. The Use Cases are: *Emotion Enhanced Speech (EES)*, *Audio Recording Preservation (ARP)*, *Speech Restoration System (SSR)*, and *Enhanced Audioconference Experience (EAE)*; the Composite AIM is *Audio Scene Description (ASD)*.

Each Use Case normatively defines:

1. The Functions of the AIW and of the AIMs.
2. The Connections between and among the AIMs.
3. The Semantics and the Formats of the input and output data of the AIW and the AIMs.

The word *normatively* implies that an Implementation claiming Conformance to:

1. An *AIW*, shall:
 - a. Have the AIW Function specified in the appropriate Section of Chapter 4.2.
 - b. Have all its AIMs and their Connections conforming with the AIW Reference Model specified in the appropriate Section of Chapter 4.2.
 - c. The AIW and AIM input and output data should have the Formats specified in the appropriate Subsection of Section 6.
2. An *AIM*, shall:
 - a. Have the AIM Function specified by the appropriate Section of Chapter 4.2.
 - b. Have input and output data Formats conforming with the appropriate Subsection of Section 6.
 - c. Receive as input and produce as output data having the Format specified in Section 6.
3. A data *Format*, the data shall have the Format specified in Section 6.

Users of this Technical Specification should note that:

1. This Technical Specification defines Interoperability Levels but does not mandate any.
2. Implementers are free to decide the Interoperability Level their Implementation should satisfy.
3. Implementers can use the Reference Software specification to develop their Implementations.
4. The Conformance Testing specification can be used to test the conformity of an Implementation to this Standard.
5. Performance Assessors can assess the level of Performance of an Implementation based on the Performance Assessment specification of this Standard.
6. Implementers and Users should consider the notices and disclaimers of Annex 2.

MPAI-CAE V2.1 includes:

1. The Scope (This Chapter)
2. Terms and Definitions
3. Normative and Informative References
4. Use Cases
5. Audio Scene Description Composite AIM
6. Data Types

7. Annexes (general concerns, some informative)

This version of the MPAI-CAE Technical Specification has been developed by the CAE-DC Development Committee. Future Versions may revise and/or extend the Scope of the Standard.

3 Terms and Definitions

The Terms used in this standard whose first letter is capital have the meaning defined in *Table 1*. The general MPAI Terms are defined in *Table 33*.

Table 1 – Table of terms and definitions

Term	Definition
Access Copy Files	Set of files providing the information stored in an audio tape recording, including Restored Audio Files, suitable for audio information access, but not for long-term preservation.
Audio	Digital representation of an analogue audio signal sampled at a frequency between 8-192 kHz with a number of bits/sample between 8 and 64.
Audio Block	A set of consecutive Audio samples.
Audio Channel	A sequence of Audio Blocks.
Audio File	A .wav file [9].
Audio Object	Audio source which is in the audible frequency band.
Audio Scene Geometry	Spatial information for the Audio Objects which are included in an audio scene.
Audio Segment	An Audio Block with Start Time and an End Time Labels corresponding to the time of the first and last sample of the Audio Segment, respectively.
Audio-Visual File	A file containing audio and video according to the MP4 File Format [13].
Capstan	The capstan is a rotating spindle used to move recording tape through the mechanism of a tape recorder.
Damaged List	A list of strings of Texts corresponding to the Damaged Segments (if any) requiring replacement with synthetic segments.
Damaged Section	An Audio Segment which is damaged in its entirety and is contained in a Damaged Segment.
Damaged Segment	An Audio Segment containing only speech (and not containing music or other sounds) which is either damaged in its entirety or contains one or more Damaged Sections specified in the Damaged List.
Degree	Strength of a feature, specifically, with respect to Emotion, “High,” “Medium,” or “Low.”
Editing List	The description of the speed, equalisation and reading backwards corrections occurred during the restoration process.
Emotion	A Data Type representing the internal status of a human or avatar resulting from their interaction with the context or subsets of it, such as “Angry”, and “Sad”.
Emotionless Speech	An Audio File containing speech without music and other sounds, and in which little or no identifiable emotion is perceptible by native listeners.
Irregularity	An event of interest to preservation in <i>Table 28</i> and <i>Table 29</i>

Irregularity File	A JSON file containing information about Irregularities of the ARP inputs.
Irregularity Image	An image corresponding to an Irregularity.
JSON	JavaScript object notation [17].
Microphone Array Geometry	Description of the position of each microphone comprising the microphone array and specific characteristics such as microphone type, look directions, and the array type.
Model Utterance	An Audio Segment used as a model or demonstration of the Emotion to be added to Emotionless Speech in order to produce Speech with Emotion.
Multichannel Audio	A data structure containing at least 2 time-aligned interleaved Audio Channels.
Multichannel Audio Stream	A data structure containing Audio Objects packaged with Audio Scene Geometry.
Neural Network Speech Model	A Neural Network Model trained on Speech Segments for Modelling and used to synthesize replacements for the entire Damaged Segment or Damaged Sections within it.
Passthrough AIM	An AIM with the same input and output data of an AIM without executing the Function of that AIM. E.g., a Noise Cancellation AIM that does not cancel the noise.
Preservation Audio File	The input Audio File resulting from the digitisation of an audio open-reel tape to be preserved and, in case, restored.
Preservation Audio-Visual File	The input Audio-Visual File produced by a camera pointed to the playback head of the magnetic tape recorder and the synchronised Audio resulting from the tape digitisation process.
Preservation Image	A Video frame extracted from Preservation Audio-Visual File.
Preservation Master Files	Set of files providing the information stored in an audio tape recording without any restoration. As soon as the original analogue recordings is no more accessible, it becomes the new item for long-term preservation.
Restored Audio Files	Set of Audio Files derived from the Preservation Audio File, where potential speed, equalisation or reading backwards errors that occurred in the digitisation process have been corrected.
Restored Speech Segment	An Audio Segment in which the entire segment has been replaced by a synthetic speech segment, or in which each Damaged Segment has been replaced by a synthetic speech segment.
Speech Features	Descriptor representing a variety of information elements incorporated in a Speech Segment, e.g., personal identity, Personal Status, additional factors such as vocal tension, creakiness, whispery quality, etc.
Speech Segments for Modelling	A set of Audio Files containing speech segments used to train the Neural Network Speech Model.
Speech With Emotion File	An Audio File containing speech with emotional features.
Spherical Coordinate System	A coordinate system where the position of a point is specified by three numbers: the radial distance of that point from a fixed origin, its polar angle measured from a fixed zenith direction, and the azimuthal angle of its orthogonal projection on a reference plane.

Spherical Grid Resolution	The maximum spherical angle between any two neighbouring sampled points on a sphere.
Text List	List of texts to be converted into speech by the Speech Synthesis for Restoration AIM.
Time Code	Number of ms from 1970-01-01T00:00:00.000 according to [7].
Time Label	A measure of time from a context-dependent zero time expressed as HH:mm:ss.SSS.
Transform Audio	A frequency representation of Audio.
Enhanced Transform Audio	Transform Audio whose samples are Enhanced Transform Audio samples.
Useful Signal	Digital signal resulting from the A/D conversion of the analogue signal recorded in an audio tape.

4 References

4.1 Normative References

This standard normatively references the following technical specifications, both from MP AI and other standard organisations:

1. MP AI; Technical Specification: Artificial Intelligence Framework (MP AI-AIF) V2.0; <https://mpai.community/standards/mpai-aif/>
2. MP AI; Technical Specification: Connected Autonomous Vehicles (MP AI-CAV) – Architecture V1.0; <https://mpai.community/standards/mpai-cav/>
3. MP AI; Technical Specification: Multimodal Conversation (MP AI-MMC) V2.1; <https://mpai.community/standards/mpai-mmc/>
4. MP AI; Technical Specification: Object and Scene Description (MP AI-OSD); <https://mpai.community/standards/mpai-osd/>
5. MP AI; Technical Specification: Portable Avatar Format (MP AI-PAF); <https://mpai.community/standards/mpai-paf/>
6. A Universally Unique IDentifier (UUID) URN Namespace; IETF RFC 4122; July 2005.
7. Date and Time on the Internet: Time Stamps; IETF RFC 3339; July 2002.
8. Universal Coded Character Set (UCS): ISO/IEC 10646; December 2020.
9. ITU-R BS.2088-1 (10/2019) - Long-form file format for the international exchange of audio programme materials with metadata.
10. ISO/IEC 14496-10; Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding.
11. ISO/IEC 23008-2; Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High Efficiency Video Coding.
12. ISO/IEC 23094-1; Information technology – General video coding – Part 1: Essential Video Coding.
13. ISO/IEC 14496-12; Information technology – Coding of audio-visual objects – Part 12: ISO base media file format.
14. ZIP format, <https://pkware.cachefly.net/webdocs/casestudies/APPNOTE.TXT>.
15. Neural Network Exchange Format; <https://www.khronos.org/registry/NEF/specs/1.0/nef-1.0.4.pdf>; Khronos.
16. Open Neural Network Exchange (ONNX) format; <https://www.ONNX.ai>.
17. The JavaScript Object Notation (JSON) Data Interchange Format; <https://datatracker.ietf.org/doc/html/rfc8259>; IETF rfc8259; December 2017.

18. BS EN 60094-1:1994, BS 6288-1: 1994, IEC 94-1:1981 - Magnetic tape sound recording and reproducing systems - Part 1: Specification for general conditions and requirements.
19. K. Bradley, IASA TC-04 Guidelines in the Production and Preservation of Digital Audio Objects: standards, recommended practices, and strategies., 2nd ed. International Association of Sound and Audiovisual Archives, (2009): 2014.
20. MPAI; The MPAI Statutes; <https://mpai.community/statutes/>
21. MPAI; The MPAI Patent Policy; <https://mpai.community/about/the-mpai-patent-policy/>.
22. Framework Licence of the Context-based Audio Enhancement Technical Specification (MPAI-CAE); <https://mpai.community/standards/mpai-cae/framework-licence/>
23. ITU-R BS.2088-1: Long-form file format for the international exchange of audio programme materials with metadata.
24. ITU-T T-81: Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines.

4.2 Informative References

The references provided here are for information purpose.

25. MPAI; Technical Specification: The governance of the MPAI Ecosystem (MPAI-GME) V1.1; <https://mpai.community/standards/mpai-gme/>
26. Framework Licence: Context-based Audio Enhancement (MPAI-CAE); <https://mpai.community/standards/mpai-cae/framework-licence/>
27. Ekman, Paul (1999), "Basic Emotions", in Dalglish, T; Power, M (eds.), Handbook of Cognition and Emotion (PDF), Sussex, UK: John Wiley & Sons.
28. B. Rafaely, Fundamentals of spherical array processing, Springer, 2018.

5 Use Cases

MPAI implements Use Cases with AI Workflows (AIW) conforming with Technical Specification: Artificial Intelligence Framework (MPAI-AIF) V2.1. Each AIW i Use Case includes:

1. Functions of the AIW
2. Reference Model of the AIW
3. I/O data of the AIW
4. Functions of AIMs
5. Web links to the AIW, AIMs, and JSON Metadata.

5.1 Emotion-Enhanced Speech (CAE-EES)

5.1.1 Functions

Speech carries information not only about its lexical content, but also about several other aspects including age, gender, identity, and **emotional state of the speaker**. Speech synthesis is evolving towards support of these aspects.

In many use cases, emotional force can usefully be added to speech which by default would be neutral or emotionless, possibly with grades of a particular emotion. For instance, in a human-machine dialogue, messages conveyed by the machine can be more effective if they carry emotions appropriately related to the emotions detected in the human speaker.

Emotion-Enhanced Speech (EES):

1. Enables a user to indicate a model utterance or an Emotion to obtain an emotionally charged version of a given utterance.

2. Converts an individual emotionless speech segment to a segment that has a specified emotion. Both input and output speech segments are contained in files. The desired emotion is expressed either as a tag belonging to a standard list of emotions or derived by extracting features from a model utterance. EES produces an output speech segment with emotion.

CAE-EES implementations can be used to create virtual agents communicating as naturally as possible, and thus improve the quality of human-machine interaction by bringing it closer to human-human interchange.

5.1.2 Reference Model

The Emotion-Enhanced Speech Reference Model depicted in Figure 2 supports two Modes or pathways enabling addition of emotional charge to an emotionless or neutral input utterance (Emotion-less Speech).

1. Along Pathway 1 (Figure 2), upper and middle left), a Model Utterance is input together with the neutral utterance Emotion-less Speech, so that features of the former can be captured and transferred to the latter.
2. Alternatively, along Pathway 2 (Figure 2), middle and lower left), neutral utterance Emotion-less Speech is input along with a specification of the desired Emotion. Speech Feature Analysis2 extracts Emotionless Speech Features from Emotionless Speech, which describe its initial state. These are sent to Emotion Feature Production, which produces (emotional) Neural Speech Features that can add the desired emotional charge to Emotionless Speech. These Neural Speech Features are sent to Neural Emotion Insertion, which combines Emotionless Speech and the (emotional) Neural Speech Features set. Speech with Emotion is then produced as output.

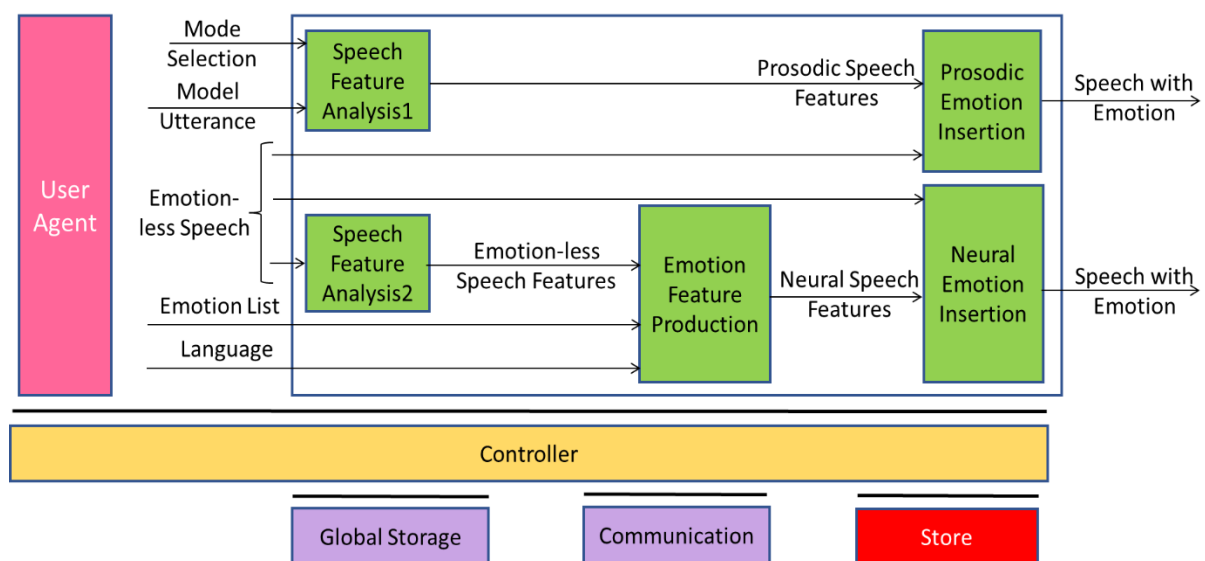


Figure 2 - Emotion-Enhanced Speech Reference Model

5.1.3 I/O data of AI Workflow

Table 2 gives the input and output data of Emotion-Enhanced Speech.

Table 2 – I/O data of Emotion-Enhanced Speech

Input data	Comments
------------	----------

Emotionless Speech	See definition in <i>Table 1</i> .
Emotion	See definition in <i>Table 1</i> .
Model Utterance	See definition in <i>Table 1</i> .
Output data	Comments
Speech with Emotion	See definition in <i>Table 1</i> .

5.1.4 Functions of AI Modules

The AI Modules perform the functions described in *Table 3*.

Table 3 – AI Modules of Emotion-Enhanced Speech

AIM	Function
Speech Feature Analysis 1	Extracts Neural Speech Features of a model emotional utterance and transfers them to the Prosodic Emotion Insertion AIM.
Speech Feature Analysis 2	Extracts Emotionless Speech Features of an emotionless input utterance, passing these to the Emotion Feature Production AIM.
Emotion Feature Production	Receives the Emotionless Speech Features produced by Speech Feature Analysis2 plus a list of Emotions to be added. (If the Degree of an Emotion is not specified, the Medium value is used.)
Prosodic Emotion Insertion	Integrates the (emotional) Prosodic Speech Features with those of the Emotionless Speech input, yielding and delivering an emotionally modified utterance.
Neural Emotion Insertion	Integrates the (emotional) Neural Speech Features with those of the Emotionless Speech input, yielding and delivering an emotionally modified utterance.

5.1.5 I/O Data of AI Modules

Table 4 – CAE-EES AIMs and their data

AIM	Input Data	Output Data
Speech Features Analysis1	Model Utterance	Prosodic Speech Features
Speech Features Analysis2	Emotionless Speech	Emotionless Speech Features
Emotion Feature Production	Emotionless Speech Features Emotion List Language	Neural Speech Features
Prosodic Emotion Insertion	Emotionless Speech Prosodic Speech Features	Speech with Emotion
Neural Emotion Insertion	Emotionless Speech Neural Speech Features	Speech with Emotion

5.1.6 AIW, AIMs, and JSON Metadata

Table 5 lists the AIW and AIMs, links to the online specifications and to the JSON Metadata.

Table 5 – AIW, AIMs, and JSON Metadata

AIW	AIMs	Name	JSON
-----	------	------	------

CAE-EES		Emotion Enhanced Speech	File
	CAE-SF1	Speech Feature Analysis 1	File
	CAE-SF2	Speech Feature Analysis 2	File
	CAE-EFP	Emotion Feature Production	File
	CAE-PEI	Prosodic Emotion Insertion	File
	CAE-NEI	Neural Emotion Insertion	File

5.2 Audio Recording Preservation (CAE-ARP)

5.2.1 Functions

Preservation of audio assets recorded on analogue media is an important activity for a variety of application domains, in particular cultural heritage. Preservation goes beyond mere A/D conversion. For instance, the magnetic tape of an open reel may hold important information: it can carry annotations (by the composer or by the technicians) or it can include multiple splices and/or display several types of Irregularities (e.g., corruptions of the carrier, tape of different colour or chemical composition). This information shall be preserved for a correct playback. Nevertheless, some errors can occur during the digitisation as well as the digitisation could be partial because of the corruption of the carrier. These errors shall be restored to make the content listenable. The ARP Use Case (see 5.1.5) concerns the creation of a digital copy of the digitized audio of open reel magnetic tapes for long-term preservation and of an access copy (restored, if necessary) for correct playback of the digitized recording.

In this Audio Recording Preservation Use Case, two files are fed into a preservation system:

1. A Preservation Audio File obtained by digitising the analogue tape audio recording composed of music, soundscape or speech read from a magnetic tape.
2. A Preservation Audio-Visual File produced by a camera pointed to the playback head of the magnetic tape recorder.

The following is not required:

1. Alignment of the start and end times of the two files. However, the maximum tolerated misalignment is 10s.
2. Presence of signal at the start and the end of the two files.
3. Alignment of the Useful Signal on both files.
4. The same time base for both files. However, the time difference between the same samples in two files shall not be more than 30ms for a 1-hour audio tape.

The output of the restoration process is composed by:

1. Preservation Master Files.
2. Access Copy Files.

5.2.2 Reference Model

Figure 3 depicts the Audio Recording Preservation Reference Model.

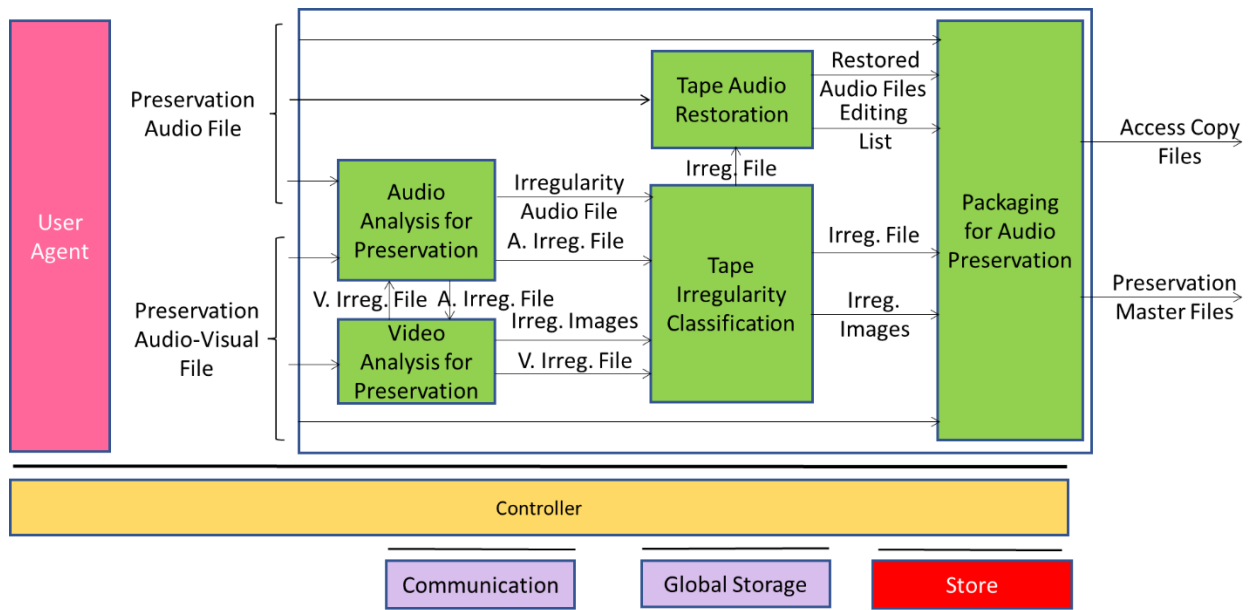


Figure 3 – Audio Recording Preservation Reference Model

The sequence of operations of the Audio Recording Preservation unfolds as follows:

1. The analogue audio signal from the open-reel tape recorder is digitised as Preservation Audio File.
2. Preservation Audio-Visual File is the combination of:
 - a. The video camera pointed at the playback head of the open-reel tape recorder.
 - b. The analogue audio signal digitised with the same video clock.
3. Audio Analysis for Preservation:
 - a. Detects Irregularities.
 - b. Assigns IDs to them that are unique to the analysed open-reel tape.
 - c. Receives an Irregularity File from the Video Analysis for Recording
 - d. Extracts the Audio Files corresponding to each Irregularity received or detected.
 - e. Sends the Audio Files and the Irregularity File related to all Irregularities to the Tape Irregularity Classification.
4. Video Analysis for Preservation:
 - a. Detects Irregularities.
 - b. Assigns IDs to them that are unique to the analysed open-reel tape.
 - c. Receives an Irregularity File from the Audio Analysis for Recording and the offset between Preservation Audio File and the Preservation Audio-Visual File.
 - d. Extracts the Irregularity Images corresponding to each Irregularity received or detected.
 - e. Sends the Irregularity Images and the Irregularity File related to all Irregularities to the Tape Irregularity Classification.
5. Tape Irregularity Classification:
 - a. Receives an Irregularity File with the corresponding Images and Audio Files.
 - b. Classifies and selects the ones considered relevant.
 - c. If the Irregularity was detected by the Video Analysis for Recording, the selected Irregularity File and the corresponding Irregularity Images are sent to the Packaging for Audio Recording.
6. The Tape Audio Restoration uses the Irregularity File to identify and restore portions of the Preservation Audio File.

7. The Packaging for Audio Preservation collects the Preservation Audio File, Restored Audio Files, the Editing List, the Irregularity File and corresponding Irregularity Images if detected by the Video Analyser, and the Preservation Audio-Visual File and then it produces the Preservation Master Files and Access Copy Files.

5.2.3 I/O data of AI Workflow

Table 6 gives the input and output data of Audio Recording Preservation.

Table 6 – I/O data of Audio Recording Preservation

Input	Comments
Preservation Audio File	A Preservation Audio File obtained by digitising the analogue tape audio recording composed of music, soundscape or speech read from a magnetic tape.
Preservation Audio-Visual File	A Preservation Audio-Visual File produced by a camera pointed to the playback head of the magnetic tape recorder.
Output data	Comments
Preservation Master Files	Set of files providing the information stored in an audio tape recording without any restoration. As soon as the original analogue recordings is no more accessible, it becomes the new item for long-term preservation.
Access Copy Files	Set of Audio Files derived from the Preservation Audio File, where potential speed, equalisation or reading backwards errors that occurred in the digitisation process have been corrected.

5.2.4 Functions of AI Modules

The AIMs required by this Use Case are described in Table 7.

Table 7 – Functions of AI Modules of Audio Recording Preservation

AIM	Function
Audio Analysis for Preservation	1. At the start, it calculates the offset between Preservation Audio and the Audio of the Preservation Audio-Visual File. 2. Sends Audio Irregularity File to and receives Video Irregularity Files from Video Analysis for Preservation. 3. Extracts the Audio Files corresponding to the Irregularities identified in both Irregularity Files. 4. Sends the Irregularity merged from the Audio and Video Irregularity Files to Tape Irregularity Classification with the corresponding Audio Files.
Video Analysis for Preservation	1. Detects and enters the Video Irregularities of the Preservation Audio-Visual File in the Video Irregularity File. 2. Sends Video Irregularity File to and receives Audio Irregularity Files from Audio Analysis for Preservation. 3. Extracts the Images corresponding to the Irregularities of both Irregularity Files.