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

MPAI Conformance Testing V1

Context-based Audio Enhancement
MPAI-CAE Version 1.4

V1

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 or any of 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 1 – Notices and Disclaimers.

Conformance Testing Specification for Context-based Audio Enhancement (MPAI-CAE) Version 1.4

1 Introduction

Technical Specification: Context-based Audio Enhancement (MPAI-CAE) [3] has been developed MPAI – Moving Picture, Audio, and Data Coding by Artificial Intelligence, the international, unaffiliated, non-profit organisation developing standards for Artificial Intelligence (AI)-based data coding with clear Intellectual Property Rights licensing frameworks [9] in compliance with the rigorous MPAI standards developing process[1] while pursuing the following policies:

1. Be friendly to the AI context but, to the extent possible, agnostic to the technology – AI or Data Processing – used in an implementation.
2. Be attractive to different industries, end users, and regulators.
3. Address three levels of standardisation: data types, components (called AI Modules), configurations of components (called AI Workflows) all exposing standard interfaces with a level of modularity decided by the implementer.
4. Specify the data exchanged by components with clear semantics to the extent possible.

As manager of the MPAI Ecosystem specified by Governance of MPAI Ecosystem (MPAI-GME) [1**Error! Reference source not found.**], MPAI also ensures that a user can:

1. Operate a reference implementation of the Technical Specification, by providing a Reference Software Specification with associated software.
2. Test the conformance of an implementation with the Technical Specification, by providing Conformance Testing Specification.
3. Assess the performance of an implementation of a Technical Specification, by providing the Performance Assessment Specification.
4. Get conforming implementations, possibly with a performance assessment report, from a trusted source through the MPAI Store.

The AI Framework (MPAI-AIF) Technical Specification [2] enables an effective implementation of some of the policies outlined above thanks to its AI Framework (AIF) environment enabling the secure execution of AI Workflows (AIW) constituted by components called AI Modules (AIM).

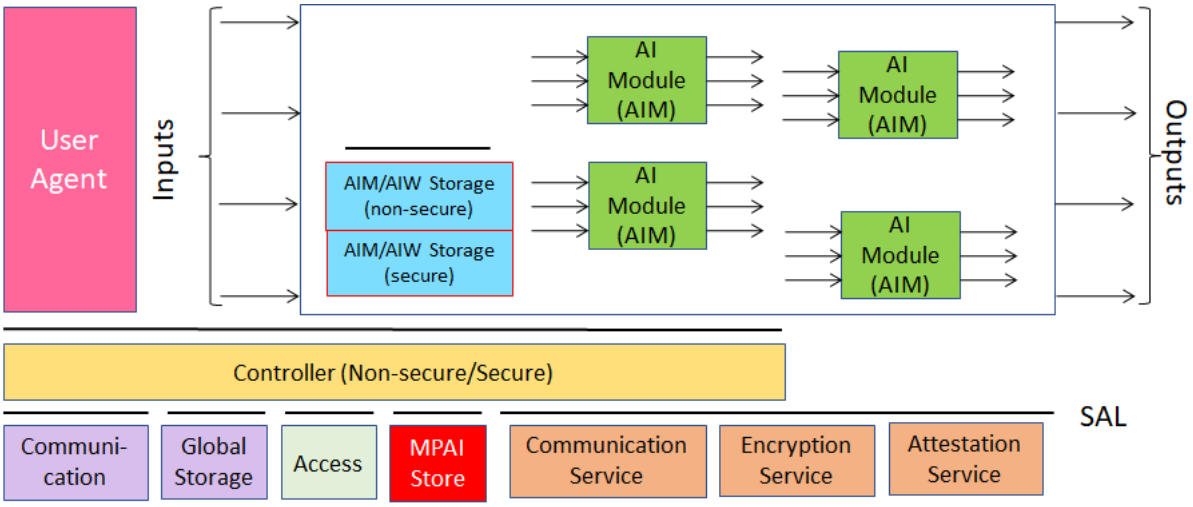

Figure 1 – The AI Framework (MPAI-AIF) V2 Reference Model

With MPAI-AIF, users can execute AI applications having an explicit computing workflow. Component developers can provide components with standard interfaces that afford improved performance compared to other implementations.

[1] https://mpai.community/about/the-mpai-patent-policy/

AIW and its AIMs may have 3 interoperability 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.
AI Modules can execute data processing or Artificial Intelligence algorithms and can be implemented in hardware, software, or hybrid hardware/software.

However, MPAI does not specify the choice of interoperability level or type, which remains the sole decision of the implementer.

This *Conformance Testing Specification: Context-based Audio Enhancement* (MPAI-CAE) specifies methods enabling users to ascertain whether a data type generated by an AIM or an AIW conform with the MPAI-CAE Technical Specification. The Version of this MPAI-CAE *Conformance Testing Specification* is Version 1.4.

The chapters and the annexes of this *Conformance Testing Specification: Context-based Audio Enhancement* (MPAI-CAE) are Normative unless they are labelled as Informative. Terms beginning with a capital letter are defined in *Table 1* if specific of this MPAI-AIF Technical Specification, or in *Table 66* is used across MPAI Standards.

2 Scope of standard

MPAI Context-based Audio Enhancement (MPAI-CAE) [3] is an MPAI Standard comprising 4 Use Cases, all sharing the characteristic of using AI to 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. This standard includes 4 Use Cases referenced as *Emotion Enhanced Speech (EES), Audio Recording Preservation (ARP), Speech Restoration System (SSR),* and *Enhanced Audioconference Experience (EAE).*
This Conformance Testing Specification (CTS) provides the MPAI Store with the Conformance Testing Dataset and/or the method to generate the Conformance Testing Dataset, the Tools, and the Procedures – in the following called the Means – to verify that the AIMs and/or the AIW of a Use Case belonging to *MPAI Context-based Audio Enhancement (MPAI-CAE)* have been properly implemented: they shall produce data whose Semantics and Format conform with the Normative clauses of the relevant Technical Specification [3] while providing at least the level of user experience specified in this document.
The *MPAI Context-based Audio Enhancement (MPAI-CAE)* Conformance Testing is based on the following process:

1. Valid Conformance Tests may only be carried out by the MPAI Store using the Means specified in this document.
2. MPAI does not make Conformance Testing Datasets publicly available for NN and AI-based AIMs.
3. In the case that the Conformance Testing Datasets are not publicly available, the MPAI Store randomly selects a large and balanced subset of test items and feeds it to the AIM under test.
4. Submitters will be informed of the outcome of the Conformance Tests through a form which is provided in this document (see Section 5).
5. Submitters will be able to verify that the AIM does indeed provide the results reported by the MPAI Store in the form:

a. By subjecting the submitted AIM to the inputs corresponding to the identifiers of the records of the Conformance Testing Dataset used by the MPAI Store
b. By verifying that the AIM produced an output having the same value as the output provided by the MPAI Store.

The number of the Conformance Testing Dataset records provided has been kept at a minimum as specified in this Conformance Testing Specification.

6. All implementers will be properly informed in two ways how the AIM has passed Conformance Tests: in a descriptive sentence and with the full table of results, as described in this document.
7. Some Conformance Testing Datasets may only be accessible within a time window and for a limited number of accesses after the test results have been received.

The current version of *MPAI Context-based Audio Enhancement (MPAI-CAE)* CTS has been developed by the *MPAI Context-based Audio Enhancement* Development Committee (CAE -DC). MPAI may decide to produce new Versions of the *MPAI Context-based Audio Enhancement (MPAI-CAE)* CTS.

3 Terms and definitions

The MPAI-specific and capitalised Terms used in this *MPAI Context-based Audio Enhancement (MPAI-CAE)* CTS have the meaning defined in *Table 1*. General MPAI Terms are defined *Table 66*.

Table 1 – MPAI-CAE Terms

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.
Accuracy	The ratio between the number of correct predictions and the total number of predictions.
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 32.
Audio Block	A set of consecutive Audio samples.
Audio Channel	A sequence of Audio Blocks.
Audio File	A .wav file.
Audio Object	Direct 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.
Capstan	The capstan is a rotating spindle used to move recording tape through the mechanism of a tape recorder.
Conformance Testing Dataset	Dataset used by the MPAI Store to Test the Conformance of an AIM to a Technical Specification.
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	One of the human emotions standardised by MPAI, or in an augmented or alternate version provided by a user.
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.
False Negative	The result of an assessment that incorrectly indicates the absence of a condition or characteristic.
False Positive	The result of an assessment that incorrectly indicates the presence of a condition or characteristic.
Interleaved Multichannel Audio	A data structure containing more than 2 time-aligned interleaved Audio Channels.
Irregularity	An event of interest to preservation.
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.
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 + Audio Scene Geometry	Multichannel Audio packaged with Audio Scene Geometry.
Neural Network Speech Model	A Neural Network Model trained on Speech Segments for Modelling and used to synthesise 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.
Power Spectral Density (PSD)	The power of a signal described as a function of its frequency components.
Precision	The ratio between the number of True Positives and the total number of True Positives and False Positives.
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.
Recall	The ratio between the number of True Positives and the total number of True Positives plus False Negatives.
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 Audio 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.
Root Mean Square Error (RMSE)	The square root of the mean square of the difference between two signals
Sensitivity	The ratio between the number of True Positives and the number of retrieved elements (i.e., True Positives and False Positives).
Specificity	The ratio between the number of True Negatives and the number of True Negatives plus False Positives.
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 Grid Resolution	The maximum spherical angle between any two neighbouring sampled points on a sphere.
Time Code	Number of ms from 1970-01-01T00:00:00.000.
Time Label	A measure of time from a context-dependent zero time expressed as HH:mm:ss.SSS.
Training Dataset	Dataset used during the training of an AIM
Transform Denoised Speech	Transform Audio whose samples are Denoised Speech samples.
True Negative	The result of an assessment that correctly indicates the absence of a condition or characteristic.
True Positive	The result of an assessment that correctly indicates the presence of a condition or characteristic.
Useful Signal	Digital signal resulting from the A/D conversion of the analogue signal recorded in an audio tape.

The acronyms used in this *MPAI Context-based Audio Enhancement (MPAI-CAE)* CTS have the meaning defined in *Table 2*.

Table 2 - List of MPAI-CAE Acronyms.

Acronym	**Meaning**
B	Brands on tape
DA	Damaged tape
DI	Dirt
EOT	Ends Of Tape
ESV	Equalization Standard Variation
M	Marks

PPS	Play, Pause and Stop
PSD	Power Spectral Density
RMSE	Root Mean Square Error
S	Shadows
SB	Signal Backward
SHD	Spherical Harmonics Decomposition
SOT	Start Of Tape
SP	Splice
SSV	Speed Standard Variation
WF	Wow and Flutter

Table 3 – Measurement metrics definitions

Metric	**Computation**
Amplitude	Given a discrete-time signal x of finite length N, the maximum displacements of the points on the signal is peak to peak Amplitude $A[n]$ and it is given as: $$A[n] = max(\Re\{x\}) - \min(\Re\{x\})$$ for each $0 \leq n < N$, where $\max(.)$ denotes the maximum value over x, and $\min(.)$ denotes the minimum value over x, where x[n] is a real or complex number.
Angle Between Two Vectors	Given two vectors $\vec{u}$ and $\vec{v}$, the angle between these two vectors is: $$\Omega = \cos^{-1}(\frac{u \cdot v}{\lvert u \rvert \lvert v \rvert})$$
Average	Given a data set $X = \{x_1, \dots, x_n\}$, containing real or complex numbers and its Average is: $$A = \frac{1}{n}\sum_{i=1}^{n} \Re\{x_i\}$$
Cross-Correlation	Given two discrete-time signals x and y of equal finite length N, the Cross-Correlation $R_{x,y}[n]$ is: $$R_{x,y}[n] = \sum_{m=0}^{N-n-1} \overline{x[m]}y[m+n]$$ for each $0 \leq n < N$, where $\overline{x[m]}$ denotes the complex conjugate of $x[m]$.
Power Spectral Density	Given a discrete-time signal x of finite length N, the Power Spectral Density of x is computed by following Welch's method [9]: 1. Partition x in K segments of equal length M. 2. Compute a windowed Discrete Fourier Transform for each segment k, $0 \leq k < K$:

	$$X_k[l] = \sum_m x_k[m]w[m]e^{-j2\pi lm/M}$$ where $0 \leq m, l < M$ and $w[m]$ is a windowing function. 3. Compute the estimate: $$S_x[l] = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{W}\lvert X_k[l]\rvert^2$$ where $W = 1/M \sum_{m=0}^{M} w^2[m]$. The parameters shall be set as follow: • $M = 4096$. • Each segment shall have $M/4$ overlapping samples. • $w[m]$ is the Hamming window.
Root Mean Squared Error	Given two discrete-time signals x and y of equal finite length N, the Root Mean Squared Error is: $$RMSE = \sqrt{\frac{\sum_{n=0}^{N-1}(\lvert x[n] - y[n]\rvert)^2}{N}}$$

4 References

4.1 Normative references

The following documents are normatively referenced by this Standard.

1. Technical Specification: The Governance of the MPAI Ecosystem (MPAI-GME) V1; https://mpai.community/standards/mpai-gme/.
2. Technical Specification: AI Framework (MPAI-AIF) V1; https://mpai.community/standards/mpai-aif/ .
3. Technical Specification MPAI Context-based Audio Enhancement (MPAI-CAE) V1.4 https://mpai.community/standards/mpai-cae/.
4. Reference Software: MPAI Context-based Audio Enhancement (MPAI- CAE) https://mpai.community/standards/mpai-cae/.
5. General licence for MPAI software; https://mpai.community/about/licence/.
6. ITU-T P.862 - Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.
7. ITU-R BS.2088-1 (10/2019) - Long-form file format for the international exchange of audio programme materials with metadata.
8. ITU-T T-81: Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines.
9. MPAI; Patent Policy; https://mpai.community/about/the-mpai-patent-policy/

4.2 Informative references

10. CAE-DC; An introduction to MPAI Context-based Audio Enhancement (MPAI-CAE); https://mpai.community/wp-content/uploads/2022/02/N549-Introduction-to-MPAI-CAE.docx.
11. P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," in IEEE Transactions on Audio and Electroacoustics, vol. 15, no. 2, pp. 70-73, June 1967, doi: 10.1109/TAU.1967.1161901.

12. Vincent, Emmanuel, Rémi Gribonval, and Cédric Févotte. "Performance measurement in blind audio source separation." IEEE transactions on audio, speech, and language processing 14.4 (2006): 1462-1469.

5 Conformance Testing

5.1 Data Sharing Procedure

In general, Conformance Testing of the CAE AIMs and AIWs requires a large set of data. The relative test results are based on scoring measures specified for each AIM/AIW. For many AIMs, the computation of a scoring measure involves a large data set to assess if the submitted implementation is conformant to the standard. When appropriate (see detailed description in the following Sections of this document), the Conformance Tester may share a small amount of data with the Submitter.

Only in case of a negative outcome, the Conformance Tester will share a small amount of data with the Submitter with the scope of verifying the testing results. The amount/type of data shared is described in the specific tables relative to the AIM under test. Typically, the shared data shall be roughly 1% of the available ground truth data used in the evaluation procedure. In addition, the shared data will be composed of two parts, one consisting of positive cases (i.e., when the test result satisfies the CT requirements) and the other of negative cases (i.e., when the test result does not satisfy the CT requirements); the two parts will always contain the same number of data sets, if possible. In case legal, ethical and privacy issues impede the sharing of specific data, no data will be provided by the Conformance Tester.

In summary, the following steps are fulfilled in the data sharing procedure.

1. The Submitter submits an AIM (or an AIW) to be verified by the Conformance Tester.
2. The Conformance Tester, following the Conformance Testing specification, assesses that the submitted AIM/AIW is *not* conformant with the standard.
3. In the Conformance Testing form the test provider shares with the Submitter the test results.
4. *If requested by the Submitter and provided for in the AIM (AIW) Conformance Testing (CT) Specification*, the Conformance Tester shares with the Submitter a small amount of data as described in the CT specification.
5. The Conformance Tester, foreseeing new submissions, marks the shared data set as "shared".

The steps above will be applied to subsequent submissions.

5.2 Emotion Enhanced Speech (EES)

The normative Architecture of Emotion Enhanced Speech (EES) is given by *Figure 2*. The figure depicts two possible paths (Modes) through the workflow, only one of which will be selected via the Mode Selection input element. The two paths represent alternative ways of adding emotion to an Emotionless Speech input segment. They could be considered as separate workflows but are described together to ease comparison and exposition. The path through Speech Feature Analyser1 assumes the use of a Model Utterance which demonstrates one way to pronounce the Emotionless Speech input segment with appropriate emotion. Speech Feature Analyser1 extracts certain speech features (Speech Features1) from this demonstration utterance for later combination with the Emotionless Speech input segment. By contrast, the path through Speech Feature Analyser2 employs Emotion List, a list of the desired emotions to be added, e.g., “angry” or “sad." Speech Feature

Analyser2 extracts certain speech features (Emotionless Speech Features) from the Emotionless Speech input segment and passes them with this list and an indication of the current Language to Emotion Feature Producer, which shall then produce additional features (Speech Features2) that can convey the specified emotions when combined with Emotionless Speech for the relevant Language. Emotion Inserter1 and Emotion Inserter2 will combine Speech Features1 or Speech Features-2 with Emotionless Speech. Note that Speech Features1 and Speech Features2 are not expected to be interchangeable.
At this stage this document does not include the specifics for some of the threshold values for this use case. The testing authority should employ widely accepted best practices.

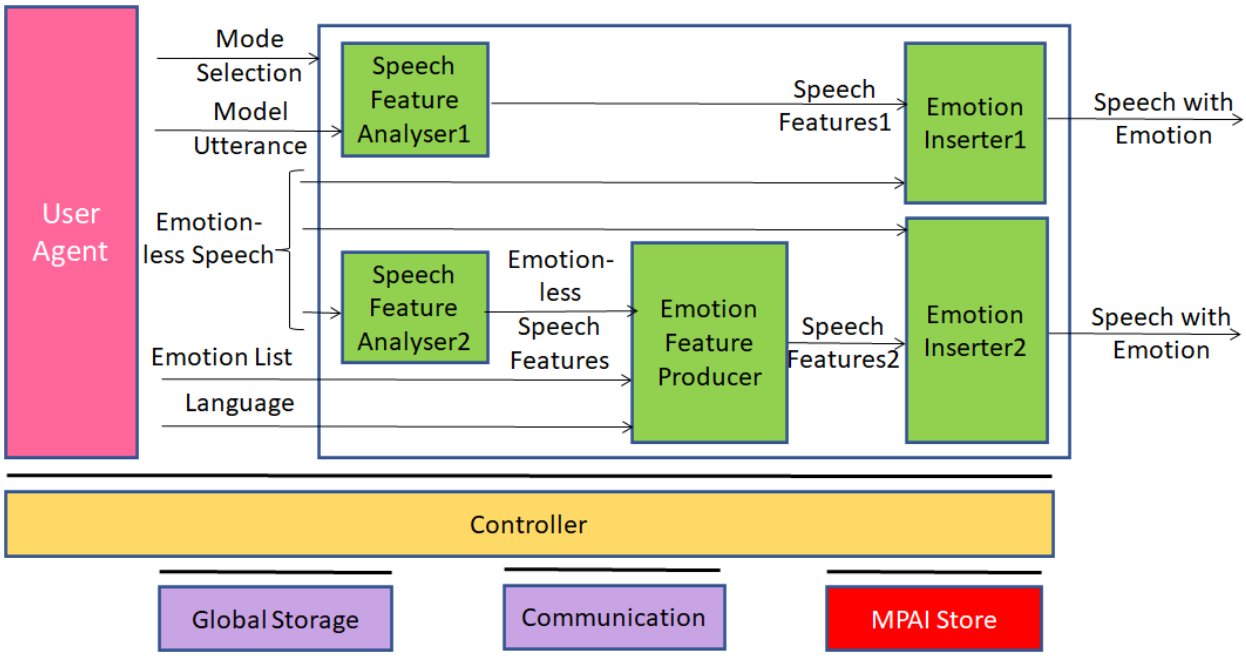

Figure 2 – Reference Model of Emotion Enhanced Speech (EES)

The input/output data format are specified in 5.1 of [3].

5.2.1 Emotion Enhanced Speech (EES) AIMs

Table 4 gives the AIMs and the input/output data of Emotion Enhanced Speech (EES).

Table 4 – AIMs and their I/O data of Emotion Enhanced Speech (EES)

AIM	Input Data	Output Data
Speech Feature Analyser 1	Model Utterance	Speech Features1
Speech Feature Analyser2	Emotionless speech	Emotionless Speech Features
Emotion Feature Producer	Emotionless Speech Features Emotion List Language	Speech Features2
Emotion Inserter1	Emotionless Speech Speech Features1	Speech with Emotion
Emotion Inserter2	Emotionless Speech Speech Features2	Speech with Emotion

5.2.2 Emotion Enhanced Speech (EES) Speech Feature Analyser1 AIM

Table 5 gives the input/output data of the Speech Feature Analyser1 AIM.

Table 5 – I/O Data of Emotion Enhanced Speech (EES) Speech Feature Analyser1 AIM

AIM	Input Data	Output Data
Speech Feature Analyser1	Model Utterance	Speech Features1

Table 6 gives the Emotion Enhanced Speech (EES) Speech Feature Analyser1 Means (verification procedures) and how they are used.

Table 6 – Means and use of Emotion Enhanced Speech (EES) Speech Feature Analyser1 AIM

Means	Actions
Conformance Testing Dataset	DS1: a dataset of at least $n > M$ Model Utterances. DS2: a dataset of n Speech Features 1 arrays, where each is associated with a specific utterance of DS1 used as input, and thus represents one correct output, given this input.
Procedure	For each of the n Model Utterances in input: 1. Feed the Speech Feature Analyser (SFA) 1 under test with the current Model Utterance. 2. Verify that the number of features in output Speech Features 1 array equals the corresponding one in DS2. 3. For each feature of the output Speech Features 1 array, compute the *delta* (absolute difference) between: a. the pitch property and the corresponding DS2 data in Hz. b. the intensity property and the corresponding DS2 data in dB. c. the duration property and the corresponding DS2 data in ms. 4. Compute the Average of: a. The *deltas* of the pitch property. b. The *deltas* of the intensity property. c. The *deltas* of the duration property. Then, compute the Average for each of the three properties among the n Model Utterances. Considering one of the three properties (pitch, intensity and duration) and denoting it as p, a mathematical representation of the computation for each property is: $$A_{p_i} = \frac{\sum_{k=1}^{m_i} \lvert SFA1_p(k) - DS2_p(k) \rvert}{m_i}$$ $$A_p = \frac{\sum_{i=1}^{n} A_{p_i}}{n}$$ Where: • For $1 \leq i \leq n$, A_{p_i} represents the Average of the *deltas* of the *i-th* Speech Features 1 array for property p. • m_i is the length of the *i-th* Speech Features 1 arrays. • For $1 \leq k \leq m_i$, $SFA1_p(k)$ and $DS2_p(k)$ denote the *k-th* value for property p of, respectively, the Speech Features 1 array coming from the Speech Feature Analyser 1 under test and the Speech Features 1 array contained in DS2.

	• A_p represents the final Average for property *p*.
Evaluation	1. Condition 2 shall be respected. 2. Given the three A_p Averages computed at the end of the Procedure, if: $$Res = \frac{A_{pitch}}{2} + \frac{A_{intensity}}{4} + \frac{A_{duration}}{4} < m$$ the Speech Feature Analyser 1 module under test has passed the Conformance Test. 3. Otherwise, Speech Feature Analyser 1 does not pass the Conformance Test.

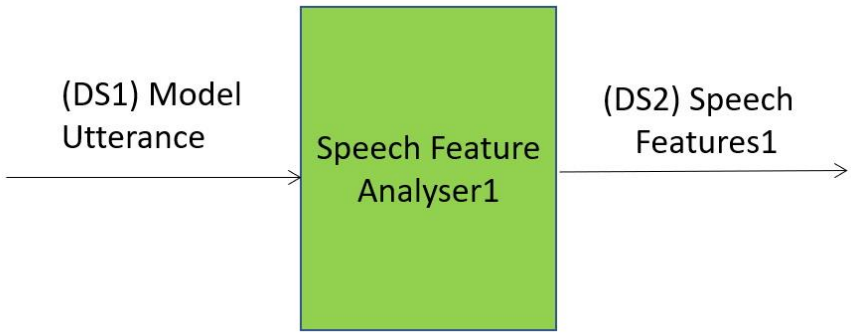

Figure 3 – EES Speech Feature Analyser1.

After the Tests, Conformance Tester shall fill out *Table 7*.

Table 7 – Conformance Testing form of Emotion Enhanced Speech (EES) Speech Feature Analyser1 (AIM1)

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:EES:1.2:0".
Name of AIM	Speech Feature Analyser1
Implementer ID	Unique Implementer Identifier assigned by Conformance Tester.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Test Dataset	Unique Dataset Identifier assigned by Conformance Tester.
Test ID	Unique Test Identifier assigned by Conformance Tester.
Actual output	Actual output provided as a matrix of *n*+1 rows containing all computed Average values: | # | Pitch | Intensity | Duration | | 1 | $A_{pitch}[1]$ | $A_{intensity}[1]$ | $A_{duration}[1]$ | | … | … | … | … | | n | $A_{pitch}[n]$ | $A_{intensity}[n]$ | $A_{duration}[n]$ | | Averages | A_{pitch} | $A_{intensity}$ | $A_{duration}$ | Result: Res Threshold: m Final evaluation: Passed / Not passed

Execution time*	Duration of test execution.
Test comment*	
Test Date	yyyy/mm/dd.

** Optional field*

5.2.3 Emotion Enhanced Speech (EES) Emotion Inserter1

Table 8 gives the input/output data of the Emotion Inserter1 AIM.

Table 8 – I/O Data of Emotion Enhanced Speech (EES) Emotion Inserter1

AIM	Input Data	Output Data
Emotion Inserter1	Emotionless Speech Speech Features1	Speech with Emotion

Table 9 gives the Emotion Enhanced Speech (EES) Emotion Inserter1 Means and how they are used.

Table 9 – AIM Means and use of Emotion Enhanced Speech (EES) Emotion Inserter1 (AIM2 in Figure 3)

Means	Actions
Conformance Testing Dataset	DS1: a dataset of at least $x > M$ Emotionless Speeches. DS2: a dataset of x Speech Features 1, each corresponding to a specific Emotionless Speech.
Procedure	For each of the x input pairs of DS1 and DS2: 1. Feed the Emotion Inserter 1 under test with an Emotionless Speech and its corresponding array of Speech Features 1. 2. Feed the reference Speech Feature Analyser 1 (ID: *S*) with the Speech with Emotion came as output from the Emotion Inserter 1 under test. 3. Verify that the number of features in Speech Features 1 array coming as output from the reference Speech Feature Analyser 1 equals the corresponding one in DS2. 4. For each feature of the output Speech Features 1 array, compute the *delta* (absolute difference) between: a. the pitch property and the corresponding DS2 data in Hz. b. the intensity property and the corresponding DS2 data in dB. c. the duration property and the corresponding DS2 data in ms. 5. Compute the Average of: a. The *deltas* of the pitch property. b. The *deltas* of the intensity property. c. The *deltas* of the duration property. Then, compute the Average for each of the three properties among the n Model Utterances.
Evaluation	1. Condition 3 shall be respected. 2. Given the three Averages computed at the end of the Procedure and denoting them with A_p, where p represents one among the three properties (pitch, intensity and duration), if:

	$$Res = \frac{A_{pitch}}{2} + \frac{A_{intensity}}{4} + \frac{A_{duration}}{4} < m$$ the Emotion Inserter 1 module under test has passed the Conformance Test. 3. Otherwise, the submitter of Emotion Inserter 1 is given the opportunity to submit an implementation of Speech Feature Analyser 1. 4. The MPAI Store will test the combination of the two submitted AIMs. 5. If the quality of the output of the submitted combination of AIM1 and AIM2 is above threshold, Emotion Inserter 1 passes the Conformance Test as long as the corresponding Speech Feature Analyser 1 is made available to the MPAI Store. 6. Else, Emotion Inserter 1 does not pass the Conformance Test.

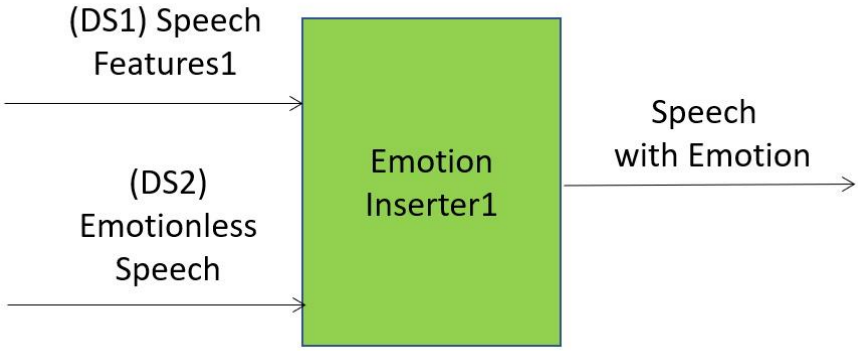

Figure 4 – EES Emotion Inserter1.

After the Tests, Conformance Tester shall fill out *Table 10*.

Table 10 – Conformance Testing form of Emotion Enhanced Speech (EES) Emotion Inserter1

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:EES:V:P".
Name of AIM	Emotion Inserter1
Implementer ID	Unique Implementer Identifier assigned by MPAI Store.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Conformance Testing Dataset	Unique Dataset Identifier assigned by MPAI Store.
Test ID	Unique Test Identifier assigned by Conformance Tester.
Actual output	Actual output provided as a matrix of n+1 rows containing all computed Average values: | # | Pitch | Intensity | Duration | | 1 | $A_{pitch}[1]$ | $A_{intensity}[1]$ | $A_{duration}[1]$ | | … | … | … | … | | n | $A_{pitch}[n]$ | $A_{intensity}[n]$ | $A_{duration}[n]$ | | Averages | A_{pitch} | $A_{intensity}$ | $A_{duration}$ | Result: Res Threshold: m

	Final evaluation: Passed / Not passed
Execution time*	Duration of test execution.
Test comment*	
Test Date	yyyy/mm/dd.

** Optional field*

5.2.4 Emotion Enhanced Speech (EES) Speech Feature Analyser2 AIM

Table 11 gives the input/output data of the Speech Feature Analyser2 AIM.

Table 11 – I/O Data of Emotion Enhanced Speech (EES) Speech Feature Analyser2 (AIM)

AIM	Input Data	Output Data
Speech Feature Analyser2	Emotionless Speech	Emotionless Speech Features

Table 12 gives the Emotion Enhanced Speech (EES) Speech Feature Analyser2 Means (verification procedures) and how they are used.

Table 12 – Means and use of Emotion Enhanced Speech (EES) Speech Feature Analyser2 AIM

Means	Actions
Conformance Testing Dataset	DS1: a dataset of at least $y > N$ Emotionless Speech Segments. DS2: a dataset of y Emotion Lists. DS3: a dataset of one element, specifying the Language in question. DS4: a dataset of y Speech with Emotion Segments, where each is associated with specific elements of DS1, DS2, and DS3 used as input, and thus represents one correct output, given this input.
Procedure	Given a reference Emotion Feature Producer (ID: *efp*), a reference Emotion Inserter 2 (ID: *ei2*) and a Speech Feature Analyser 2 module that we want to test, we measure the quality of Speech Feature Analyser 2 in relation to the reference modules as follows: 1. Connect the three modules. 2. Repeat many times: a. Select an input set comprised of a DS1 (Emotionless Speech segment), a DS2 (an Emotion List), and a DS3 (a Language). b. Feed that set to the system composed by the connected modules. c. Measure the quality of the Speech with Emotion output generated by the system by comparing it with the corresponding "correct" result in DS4 as measured with PESQ [6]. 3. The quality of Speech Feature Analyser 2 is then the *average value* of the multiple quality measurements of 2c.
Evaluation	1. If the *average value* of the quality measurements is above a threshold greater than 2.0 as specified by PESQ, Speech Feature Analyser 2 has passed the Conformance Test. 2. If the quality is below threshold, the submitter of Speech Feature Analyser 2 is given the opportunity to submit an implementation of Emotion Feature Producer and Emotion Inserter 2. 3. The MPAI Store will test the combination of the three submitted AIMs.

	4. If the quality of the output of the submitted combination is above threshold, Speech Feature Analyser 2 passes the Conformance Test as long as the corresponding Emotion Feature Producer and Emotion Inserter 2 are made available to the MPAI Store. 5. Else, Speech Feature Analyser 2 doesn't pass the Conformance Test.

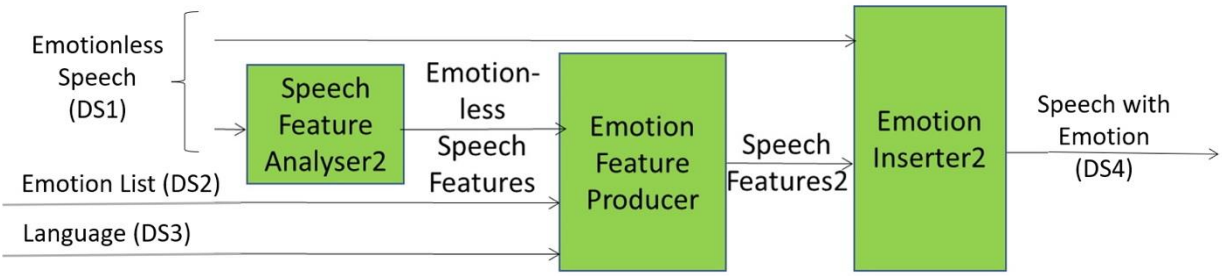

Figure 5 - EES path 2.

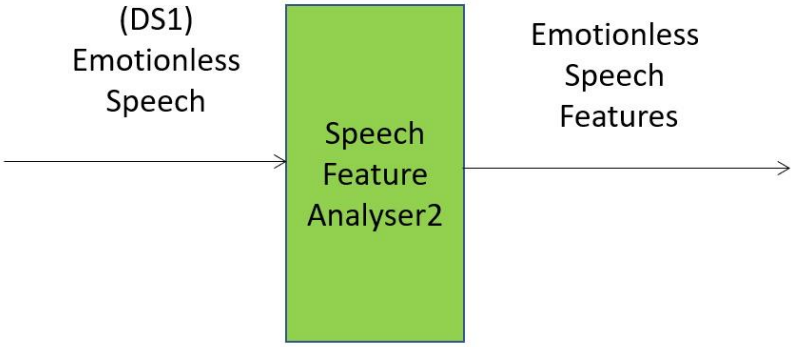

Figure 6 - EES Speech Feature Analyser2.

After the Tests, Conformance Tester shall fill out *Table 13*.

Table 13 – Conformance Testing form of Emotion Enhanced Speech (EES) Speech Feature Analyser2 AIM

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:EES:1:0".
Name of AIM	Speech Feature Analyser2
Implementer ID	Unique Implementer Identifier assigned by MPAI Store.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Conformance Testing Dataset	Unique Dataset Identifier assigned by MPAI Store.
Test ID	Unique Test Identifier assigned by Conformance Tester.
Actual output	The Conformance Tester will provide the following matrix related to the modules utilized for the tests. Denoting with i and j, $0 \leq i < x$ and $0 \leq j < y$, the record number in DS1 and DS2 respectively, the matrices reflect the results obtained with a limited number of random $[i, j]$ multiple inputs

	and the corresponding outputs. Example: DS1 | DS2 | DS4 | Emotion Inserter2 output value DS1[*i*] | DS2[*j*] | DS4[*i*, *j*] | SpeechWithEmotion[*i*, *j*] Language: DS3
Execution time*	Duration of test execution.
Test comment*	In case step 1 of Conformance Testing fails, the Conformance Tester shall request the implementer to provide an Emotion Feature Producer AIM (AIM2). In case step 4 or 5 of Conformance Testing also fails, the Conformance Tester shall inform the implementer that the Speech Feature Analyser2 (AIM1) did not pass the CT.
Test Date	yyyy/mm/dd.

** Optional field*

5.2.5 Emotion Enhanced Speech (EES) Emotion Feature Producer

Table 14 gives the input/output data of the Emotion Feature Producer AIM.

Table 14 – I/O Data of Emotion Enhanced Speech (EES) Emotion Feature Producer AIM

AIM	Input Data	Output Data
Emotion Feature Producer	Emotionless Speech Features Emotion List Language	Speech Features2

Table 15 gives the Emotion Enhanced Speech (EES) Emotion Feature Producer AIM Means and how they are used.

Table 15 – Means and use of Emotion Enhanced Speech (EES) Emotion Feature Producer AIM

Means	Actions
Conformance Testing Dataset	DS1: a dataset of at least $y > N$ Emotionless Speech Segments. DS2: a dataset of y Emotion Lists. DS3: a dataset of one element, specifying the Language in question. DS4: a dataset of y Speech with Emotion Segments, where each is associated with specific elements of DS1, DS2, and DS3 used as input, and thus represents one correct output, given this input.
Procedure	Given a reference Speech Feature Analyser 2 (ID: *sfa2*), a reference Emotion Inserter 2 (ID: *ei2*) and an Emotion Feature Producer module that we want to test, we measure the quality of Emotion Feature Producer in relation to the reference modules as follows: 4. Connect the three modules. 5. Repeat many times: a. Select an input set comprised of a DS1 (Emotionless Speech segment), a DS2 (an Emotion List), and a DS3 (a Language). b. Feed that set to the system composed by the connected modules.

	c. Verify that the Speech with Emotion output generated by the system is sufficiently close to the corresponding ground truth DS4 as specified by the threshold below. The quality of Emotion Feature Producer is then the *average value* of the multiple quality measurements of 2c.
Evaluation	1. If the *average value* of the quality measurements is above a specified threshold, Emotion Feature Producer has passed the Conformance Test. 2. If the quality is below threshold, the submitter of Emotion Feature Producer is given the opportunity to submit an implementation of Speech Feature Analyser 2 and Emotion Inserter 2. 3. The MPAI Store will test the combination of the three submitted AIMs. 4. If the quality of the output of the submitted combination is above threshold, Emotion Feature Producer passes the Conformance Test as long as the corresponding Speech Feature Analyser 2 and Emotion Inserter 2 are made available to the MPAI Store. 5. Else, Emotion Feature Producer doesn't pass the Conformance Test.

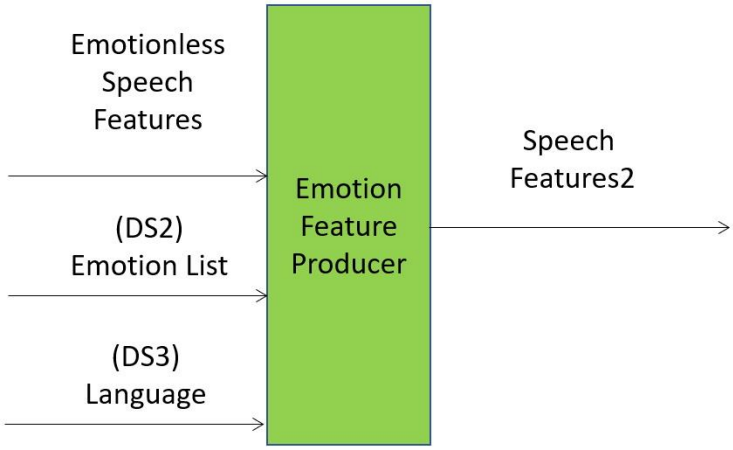

Figure 7 – Emotion Feature Producer.

After the Tests, Conformance Tester shall fill out *Table 16*.

Table 16 – Conformance Testing form of Emotion Enhanced Speech (EES) Emotion Feature Producer

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:EES:1:0".
Name of AIM	Emotion Feature Inserter
Implementer ID	Unique Implementer Identifier assigned by MPAI Store.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Conformance Testing Dataset	Unique Dataset Identifier assigned by MPAI Store.
Test ID	Unique Test Identifier assigned by Conformance Tester.
Actual output	The Conformance Tester will provide the following matrix related to the modules utilized for the tests. Denoting with *i*

	and j, $0 \leq i < x$ and $0 \leq j < y$, the record number in DS1 and DS2 respectively, the matrices reflect the results obtained with a limited number of random $[i, j]$ multiple inputs and the corresponding outputs. Example: DS1 | DS2 | DS4 | Emotion Inserter2 output value DS1[i] | DS2[j] | DS4[i, j] | SpeechWithEmotion[i, j] Language: DS3
Execution time*	Duration of test execution.
Test comment*	In case step 1 of Conformance Testing fails, the Conformance Tester shall request the implementer to provide a Speech Feature Analyser2 and Emotion Inserter2 AIMs. In case step 4 or 5 of Conformance Testing also fails, the Conformance Tester shall inform the implementer that the Emotion Feature Producer did not pass the CT.
Test Date	yyyy/mm/dd.

** Optional field*

5.2.6 Emotion Enhanced Speech (EES) Emotion Inserter2

Table 17 gives the input/output data of the Emotion Inserter2 AIM.

Table 17 – I/O Data of Emotion Enhanced Speech (EES) Emotion Inserter2

AIM	Input Data	Output Data
Emotion Inserter2	Emotionless Speech Speech Features2	Speech with Emotion

Table 18 gives the Emotion Enhanced Speech (EES) Emotion Inserter2 Means and how they are used.

Table 18 – AIM Means and use of Emotion Enhanced Speech (EES) Emotion Inserter2

Means	Actions
Conformance Testing Dataset	DS1: a dataset of at least $y > N$ Emotionless Speech Segments. DS2: a dataset of y Emotion Lists. DS3: a dataset of one element, specifying the Language in question. DS4: a dataset of y Speech with Emotion Segments, where each is associated with specific elements of DS1, DS2, and DS3 used as input, and thus represents one correct output, given this input.
Procedure	Given a reference Speech Feature Analyser 2 (ID: *sfa2*), a reference Emotion Feature Producer (ID: *efp*) and an Emotion Inserter 2 module that we want to test, we measure the quality of Emotion Inserter 2 in relation to the reference modules as follows: 1. Connect the three modules. 2. Repeat many times: a. Select an input set comprised of a DS1 (Emotionless Speech segment), a DS2 (an Emotion List), and a DS3 (a Language). b. Feed that set to the system composed by the connected modules.

	c. Measure the quality of the Speech with Emotion output generated by the system by comparing it with the corresponding "correct" result in DS4 as measured by PESQ [6]. 3. The quality of Emotion Inserter 2 is then the *average value* of the multiple quality measurements of 2c.
Evaluation	1. If the *average value* of the quality measurements is above a threshold above 2.0 as specified by PESQ, Emotion Inserter 2 has passed the Conformance Test. 2. If the quality is below threshold, the submitter of Emotion Inserter 2 is given the opportunity to submit an implementation of Speech Feature Analyser 2 and Emotion Feature Producer. 3. The MPAI Store will test the combination of the three submitted AIMs. 4. If the quality of the output of the submitted combination is above threshold, Emotion Inserter 2 passes the Conformance Test as long as the corresponding Speech Feature Analyser 2 and Emotion Feature Producer are made available to the MPAI Store. 5. Else, Emotion Inserter 2 doesn't pass the Conformance Test.

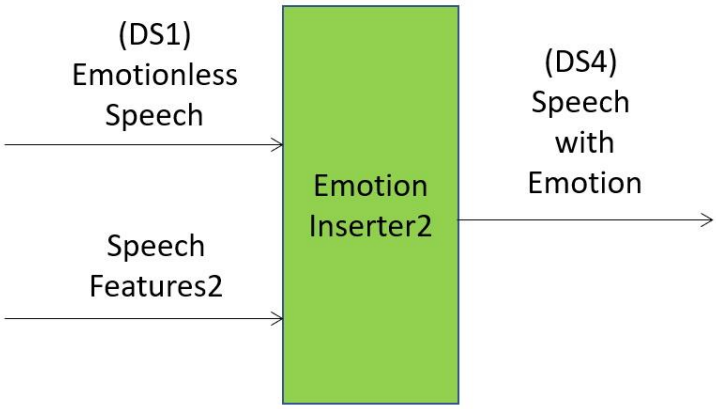

Figure 8 – *Emotion Inserter2.*

After the Tests, Conformance Tester shall fill out *Table 19*.

Table 19 – Conformance Testing form of Emotion Enhanced Speech (EES) Emotion Inserter2

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:EES:V:P".
Name of AIM	Emotion Inserter2
Implementer ID	Unique Implementer Identifier assigned by MPAI Store.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Conformance Testing Dataset	Unique Dataset Identifier assigned by MPAI Store.
Test ID	Unique Test Identifier assigned by Conformance Tester.
Actual output	The Conformance Tester will provide the following matrix related to the modules utilized for the tests. Denoting with *i*

<table>
<tr><td></td><td>and j, $0 \leq i < x$ and $0 \leq j < y$, the record number in DS1 and DS2 respectively, the matrices reflect the results obtained with a limited number of random $[i, j]$ multiple inputs and the corresponding outputs.
Example:
<table>
<tr><td>DS1</td><td>DS2</td><td>DS4</td><td>Emotion Inserter2 output value</td></tr>
<tr><td>DS1[i]</td><td>DS2[j]</td><td>DS4[i, j]</td><td>SpeechWithEmotion[i, j]</td></tr>
</table>
Language: DS3</td></tr>
<tr><td>Execution time*</td><td>Duration of test execution.</td></tr>
<tr><td>Test comment*</td><td>In case step 1 of Conformance Testing fails, the Conformance Tester shall request the implementer to provide a Speech Feature Analyser2 and Emotion Feature Producer AIMs.
In case step 4 or 5 of Conformance Testing also fails, the Conformance Tester shall inform the implementer that the Emotion Inserter2 did not pass the CT.</td></tr>
<tr><td>Test Date</td><td>yyyy/mm/dd.</td></tr>
</table>

** Optional field*

5.3 Audio Recording Preservation (ARP)

The normative Architecture of the Audio Recording Preservation (CAE-ARP) is given by *Figure 9*.

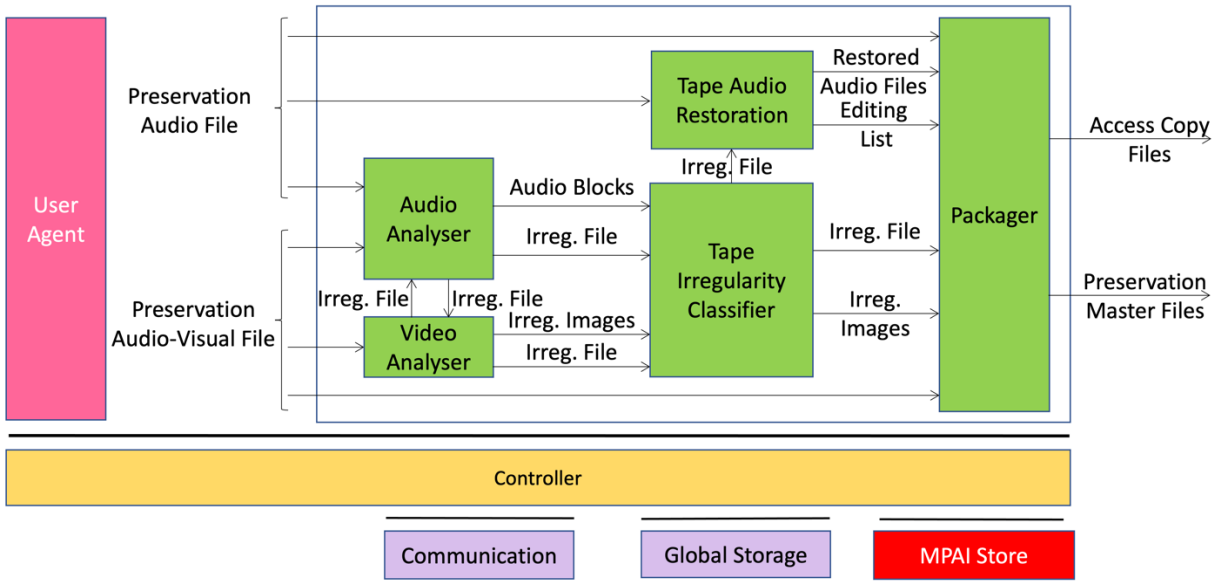

Figure 9 – Reference Model of Audio Recording Preservation (CAE-ARP).

The input/output data format are specified in 5.1.2 [3].

Table 20 shows the acronyms of ARP.

Table 20 – List of ARP acronyms.

Acronym	Meaning
B	Brands on tape

DA	Damaged tape
DI	Dirt
EOT	Ends Of Tape
ESV	Equalization Standard Variation
M	Marks
PPS	Play, Pause and Stop
PSD	Power Spectral Density
RMSE	Root Mean Square Error
S	Shadows
SB	Signal Backward
SOT	Start Of Tape
SP	Splice
SSV	Speed Standard Variation
WF	Wow and Flutter

5.3.1 Audio Recording Preservation (ARP) AIMs

Table *21* gives the AIMs and the input/output data of Audio Recording Preservation (ARP).

Table 21 *– AIMs and their I/O data of Audio Recording Preservation (ARP).*

AIM	Input Data	Output Data
Audio Analyser	Preservation Audio File Preservation Audio-Visual File Irregularity File	Audio Files Irregularity File
Video analyser	Preservation Audio-Visual File Irregularity File	Irregularity File Irregularity Images
Tape Irregularity classifier	Audio Files Irregularity Images Irregularity File	Irregularity File Irregularity Images
Tape Audio Restoration	Irregularity File Preservation Audio File	Editing List Restored Audio Files
Packager	Preservation Audio File Restored Audio Files Editing List Irregularity File Irregularity Images Preservation Audio-Visual File	Access Copy Files Preservation Master Files

5.3.2 Audio Recording Preservation (ARP) Audio Analyser

Table 22 gives the input/output data of the Audio Analyser.

Table 22 – I/O Data of Audio Recording Preservation (ARP) Audio Analyser.

AIM	Input Data	Output Data
Audio Analyser	Preservation Audio File Preservation Audio-Visual File Irregularity File	Audio Files Irregularity File

Table 23 gives the Audio Recording Preservation (ARP) Audio Analyser Means and how they are used.

Table 23 – AIM Means and use of Audio Recording Preservation (ARP) Audio Analyser.

Means	**Actions**
Conformance Testing Dataset	DS1: *n** Preservation Audio Files. DS2: *n* Preservation Audio-Visual Files related to DS1. DS3: *n* Irregularity Files related to DS2. DS4: *n* output Irregularity Files in the format of port IrregularityFileOutput_1 with all Irregularities correctly identified. DS5: *n* output Irregularity Files in the format of port IrregularityFileOutput_2 with the real offset and all Irregularities correctly identified and included from DS3. * A reasonable n for testing is 5<n<=10, since each file generates multiple irregularities to classify
Procedure	1. Feed Audio Analyser under test with DS1, DS2 and DS3. 2. Compare the computed offsets with the ones contained in DS5. 3. Analyse the Irregularity Files resulting from port IrregularityFileOutput_1. 4. Analyse the Irregularity Files resulting from port IrregularityFileOutput_2.
Evaluation	1. Verify the conditions: a. The Irregularity Files are syntactically correct and conforming to the JSON schema provided in CAE Technical Specification. b. All Irregularities from DS3 are included in the Irregularity Files coming from port IrregularityFileOutput_2. c. $\lvert O_c - O_r \rvert < 3 \times \left\lceil \frac{1000}{FPS_{DS3}} \right\rceil ms$, where O_c is the offset computed by the Audio Analyser under test, O_r is the real offset and FPS_{DS3} is the number of frames per second at which the DS3 video has been recorded. d. All output Audio Files are conforming to RF64 file format [7]. e. For each of the *n* tuples of input records, the output Audio Files are extracted from the corresponding input Preservation Audio File at the Time Labels indicated in the Irregularity File coming from port IrregularityFileOutput_2. 2. By inspecting the Irregularity Files resulting from port IrregularityFileOutput_1, for each of the *n* tuples of input records, compute the values of Recall (R) and Precision (P). 3. Compute the average value of Recall ($\tilde{R}$) and Precision ($\tilde{P}$) measures obtained at point 2. 4. Accept the AIM under test if: a. $\tilde{R} > 0.9$ b. $\tilde{P} > 0.9$

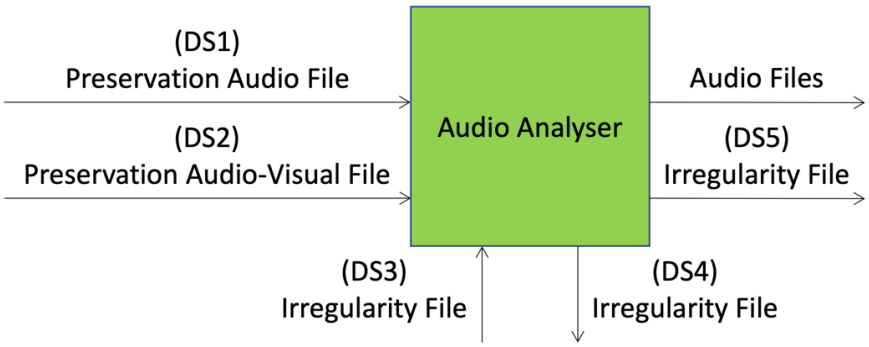

Figure 10 – *Audio Analyser.*

After the Tests, Conformance Tester shall fill out Table 24.

Table 24 – Conformance Testing form of Audio Recording Preservation (ARP) Audio Analyser.

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:ARP:1:0".
Name of AIM	Audio Analyser
Implementer ID	Unique Implementer Identifier assigned by MPAI Store.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Conformance Testing Dataset	Unique Dataset Identifier assigned by MPAI Store.
Test ID	Unique Test Identifier assigned by Conformance Tester.
Actual output	Actual output provided as a matrix of *n* rows containing *R* and *P* values.

Tuple #	*R*	*P*
1	Measure 1	Measure 1
…	…	…
n	Measure *n*	Measure *n*

Execution time*	Duration of test execution.
Test comment*	-
Test Date	yyyy/mm/dd.

** Optional field*

5.3.3 Audio Recording Preservation (ARP) Video Analyser

Table 25 gives the input/output data of the Video Analyser.

Table 25 – I/O Data of Audio Recording Preservation (ARP) Video Analyser.

AIM	Input Data	Output Data
Video Analyser	Preservation Audio-Visual File Irregularity File	Irregularity File Irregularity Images

Table 26 gives the Audio Recording Preservation (ARP) *Video Analyser* Means and how they are used.

Table 26 – AIM Means and use of Audio Recording Preservation (ARP) Video Analyser.

Means	**Actions**
Conformance Testing Dataset	DS1: *n* Preservation Audio-Visual Files. DS2: *n* Irregularity Files related to the Preservation Audio File related to DS1. DS3: *n* output Irregularity Files in the format of port IrregularityFileOutput_1 with all Irregularities correctly identified. DS4: *n* output Irregularity Files in the format of port IrregularityFileOutput_2 with all Irregularities correctly identified and included from DS3.
Procedure	1. Feed Video Analyser under test with DS1 and DS2. 2. Analyse the Irregularity Files resulting from port IrregularityFileOutput_1. 3. Analyse the Irregularity Files resulting from port IrregularityFileOutput_2.
Evaluation	1. Verify the conditions: a. The Irregularity Files are syntactically correct and conforming to the JSON schema provided in CAE Technical Specification. b. All Irregularities from DS2 are included in the Irregularity Files resulting from port IrregularityFileOutput_2. c. All output Irregularity Images are conforming to the JPEG standard [8]. d. For each of the *n* tuples of input records, the output Irregularity Images are extracted from the corresponding input Preservation Audio-Visual File at the Time Labels indicated in the Irregularity Files coming from port IrregularityFileOutput_2. 2. By inspecting the Irregularity Files resulting from port IrregularityFileOutput_1, for each of the *n* tuples of input records, compute the values of Recall (R) and Precision (P). 3. Compute the average value of Recall ($\tilde{R}$) and Precision ($\tilde{P}$) measures obtained at point 2. 4. Accept the AIM under test if: a. $\tilde{R} > 0.9$ b. $\tilde{P} > 0.9$

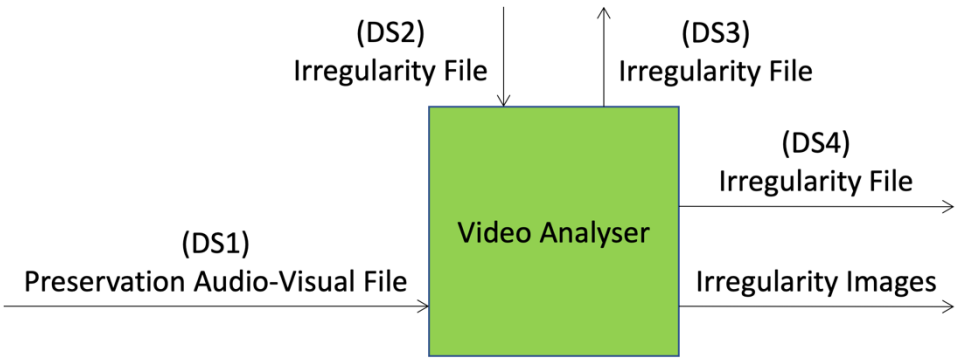

Figure 11 – *Video Analyser.*

After the Tests, Conformance Tester shall fill out Table 27.

Table 27 – Conformance Testing form of Audio Recording Preservation (ARP) Video Analyser.

<table>
<tr><td>Conformance Tester ID</td><td>Unique Conformance Tester Identifier assigned by MPAI</td></tr>
<tr><td>Standard, Use Case ID and Version</td><td>Standard ID and Use Case ID, Version and Profile of the standard in the form “CAE:ARP:1:0”.</td></tr>
<tr><td>Name of AIM</td><td>Video Analyser</td></tr>
<tr><td>Implementer ID</td><td>Unique Implementer Identifier assigned by MPAI Store.</td></tr>
<tr><td>AIM Implementation Version</td><td>Unique Implementation Identifier assigned by Implementer.</td></tr>
<tr><td>Neural Network Version*</td><td>Unique Neural Network Identifier assigned by Implementer.</td></tr>
<tr><td>Identifier of Conformance Testing Dataset</td><td>Unique Dataset Identifier assigned by MPAI Store.</td></tr>
<tr><td>Test ID</td><td>Unique Test Identifier assigned by Conformance Tester.</td></tr>
<tr><td>Actual output</td><td>Actual output provided as a matrix of n rows containing R and P values.
<table>
<tr><td>Tuple #</td><td>R</td><td>P</td></tr>
<tr><td>1</td><td>Measure 1</td><td>Measure 1</td></tr>
<tr><td>…</td><td>…</td><td>…</td></tr>
<tr><td>n</td><td>Measure n</td><td>Measure n</td></tr>
</table></td></tr>
<tr><td>Execution time*</td><td>Duration of test execution.</td></tr>
<tr><td>Test comment*</td><td>-</td></tr>
<tr><td>Test Date</td><td>yyyy/mm/dd.</td></tr>
</table>

** Optional field*

5.3.4 Audio Recording Preservation (ARP) Tape Irregularity Classifier

Table 28 gives the input/output data of the Tape Irregularity Classifier.

Table 28 – I/O Data of Audio Recording Preservation (ARP) Tape Irregularity Classifier.

AIM	Input Data	Output Data
Tape Irregularity Classifier	Audio Files Irregularity Images Irregularity File	Irregularity File Irregularity Images

Table 29 gives the Audio Recording Preservation (ARP) *Tape Irregularity Classifier* Means and how they are used.

Table 29 – AIM Means and use of Audio Recording Preservation (ARP) Tape Irregularity Classifier.

Means	Actions
Conformance Testing Dataset	DS1: *n* Irregularity Files from Audio Analyser. DS2: *n* Audio Files related to DS1. DS3: *n* Irregularity Files from Video Analyser. DS4: *n* Irregularity Images related to DS3. DS5: *n* output Irregularity Files in the format of port IrregularityFileOutput_1, containing correctly classified Irregularities. DS6: *n* output Irregularity Files in the format of port IrregularityFileOutput_2, containing correctly classified Irregularities.

Procedure	1. Feed Tape Irregularity Classifier under test with DS1, DS2, DS3 and DS4. 2. Analyse the Irregularity Files resulting from port IrregularityFileOutput_1. 3. Analyse the Irregularity Files resulting from port IrregularityFileOutput_2.
Evaluation	1. Verify the conditions: a. The Irregularity Files are syntactically correct and conforming to the JSON schema provided in CAE Technical Specification. b. The Irregularity Files resulting from port IrregularityFileOutput_1 contain only Irregularities of interest for the Tape Audio Restoration (i.e., Irregularities with IrregularityType SSV, ESV or SB). c. All output Irregularity Images are conforming to the JPEG standard [8]. d. For each of the *n* tuples of input records, the output Irregularity Images are equal to the input Irregularity Images corresponding to the Time Labels indicated in the Irregularity Files coming from port IrregularityFileOutput_2. 2. By inspecting the Irregularity Files resulting from port IrregularityFileOutput_1, for each of the *n* tuples of input records, compute the values of Recall (R) and Precision (P) for each of the 13 labels of IrregularityType defined in Tables 17 and 18 of [3]. 3. For each label l of IrregularityType, compute the average value of Recall ($\hat{R}$) and Precision ($\hat{P}$) measures obtained at point 2. 4. Compute the *average value* of Recall ($\tilde{R}$) and Precision ($\tilde{P}$) measures obtained at point 3. 5. Accept the AIM under test if: a. $\tilde{R} > 0.9$ b. $\tilde{P} > 0.9$

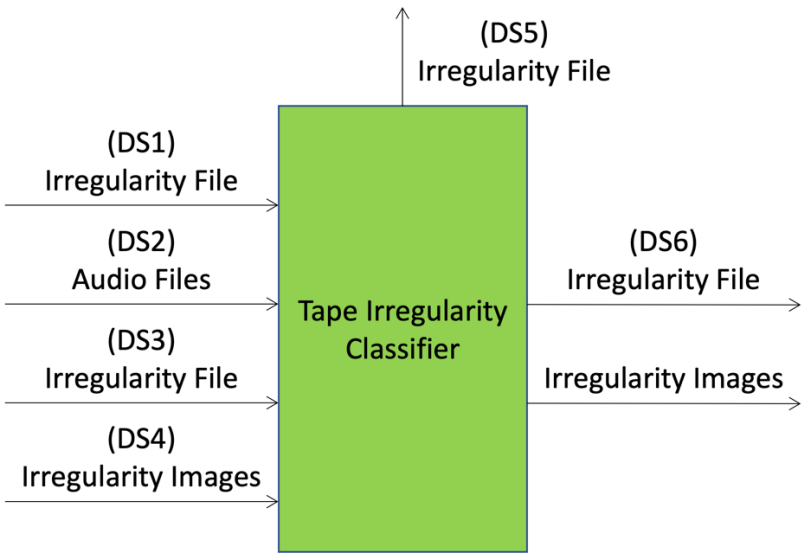

Figure 12 – *Tape Irregularity Classifier.*

After the Tests, Conformance Tester shall fill out Table 30.

Table 30 – Conformance Testing form of Audio Recording Preservation (ARP) Tape Irregularity Classifier.

<table>
<tr><td>Conformance Tester ID</td><td>Unique Conformance Tester Identifier assigned by MPAI</td></tr>
<tr><td>Standard, Use Case ID and Version</td><td>Standard ID and Use Case ID, Version and Profile of the standard in the form “CAE:ARP:1:0”.</td></tr>
<tr><td>Name of AIM</td><td>Tape Irregularity Classifier</td></tr>
<tr><td>Implementer ID</td><td>Unique Implementer Identifier assigned by MPAI Store.</td></tr>
<tr><td>AIM Implementation Version</td><td>Unique Implementation Identifier assigned by Implementer.</td></tr>
<tr><td>Neural Network Version*</td><td>Unique Neural Network Identifier assigned by Implementer.</td></tr>
<tr><td>Identifier of Conformance Testing Dataset</td><td>Unique Dataset Identifier assigned by MPAI Store.</td></tr>
<tr><td>Test ID</td><td>Unique Test Identifier assigned by Conformance Tester.</td></tr>
<tr><td>Actual output</td><td>Actual output provided as a matrix of n rows containing R and P values.
<table>
<tr><td>Tuple #</td><td>Label</td><td>R</td><td>P</td></tr>
<tr><td>1</td><td>SP</td><td>Measure 1</td><td>Measure 1</td></tr>
<tr><td>…</td><td>…</td><td>…</td><td>…</td></tr>
<tr><td>1</td><td>SB</td><td>Measure l</td><td>Measure l</td></tr>
<tr><td>…</td><td>…</td><td>…</td><td>…</td></tr>
<tr><td>n</td><td>SP</td><td>Measure $(n-1)*l+1$</td><td>Measure $(n-1)*l+1$</td></tr>
<tr><td>…</td><td>…</td><td>…</td><td>…</td></tr>
<tr><td>n</td><td>SB</td><td>Measure $n*l$</td><td>Measure $n*l$</td></tr>
</table></td></tr>
<tr><td>Execution time*</td><td>Duration of test execution.</td></tr>
<tr><td>Test comment*</td><td>-</td></tr>
<tr><td>Test Date</td><td>yyyy/mm/dd.</td></tr>
</table>

** Optional field*

5.3.5 Audio Recording Preservation (ARP) Tape Audio Restoration

Table 31 gives the input/output data of the Tape Audio Restoration.

Table 31 – I/O Data of Audio Recording Preservation (ARP) Tape Audio Restoration.

AIM	Input Data	Output Data
Tape Audio Restoration	Irregularity File Preservation Audio File	Editing List Restored Audio Files

Table 32 gives the Audio Recording Preservation (ARP) *Tape Audio Restoration* Means and how they are used.

Table 32 – AIM Means and use of Audio Recording Preservation (ARP) Tape Audio Restoration.

Means	Actions
Conformance Testing Dataset	DS1: *n* Preservation Audio Files created with (i) SB, (ii) SSV or (iii) ESV errors. These kinds of errors can occur singularly or superimposed one over another. DS2: *n* Irregularity Files related to DS1 coming from Tape Irregularity Classifier.

	DS3: *n* Restored Audio Files arrays containing the corrected files generated from DS1 with the information contained in DS2. DS4: *n* Editing Lists JSONs containing the edits made in DS3.
Procedure	1. Feed Tape Audio Restoration under test with DS1 and DS2. 2. Compare the output Editing Lists with DS4. 3. Compare the samples of output Restored Audio Files with DS3 for SB and SSV correction evaluation. 4. Compare the samples and the spectral content of output Restored Audio Files with DS3 for ESV correction evaluation.
Evaluation	1. Verify the conditions: a. The Editing Lists are syntactically correct and conforming to the JSON schema provided in CAE Technical Specification. b. The output Editing Lists contain the same edits listed in DS4. c. All output Restored Audio Files are conforming to RF64 file format [7]. 2. Whenever SB or SSV corrections are required, each file of the output Restored Audio Files array: a. Shall be of the same duration of the corresponding file in DS3. b. Shall present the maximum value of the Cross-Correlation function for $i = 0$. Considering the Cross-Correlation definition in **Error! Reference source not found.**, $x(i)$ is the Restored Audio File under evaluation and $y(i)$ is the corresponding file in DS3. 3. Whenever an ESV correction is <u>required</u>, each file of the output Restored Audio Files shall have: a. Time domain samples with RMSE < *0.1*A*. Considering the RMSE definition in **Error! Reference source not found.**, $x(i)$ is the Restored Audio File under evaluation and $y(i)$ is the corresponding file in DS3. Where A is the Amplitude of the signal from DS3. b. $\lvert S_{AIM}(f) - S_{DS3}(f)\rvert < \lvert S_{AIM}(f) - S_{DS1}(f)\rvert$ for f in [20, 20k] Hz, where $S_{AIM}(f)$ is the PSD of the AIM (Tape Audio Restoration) under test, $S_{DS3}(f)$ is the PSD of the corresponding file in DS3 and $S_{DS1}(f)$ is the PSD of the corresponding Preservation Audio File Audio Segment. 4. If, for any of the *n* input tuples, the above conditions are not satisfied, the Tape Audio Restoration module under test does not pass the Conformance Test.

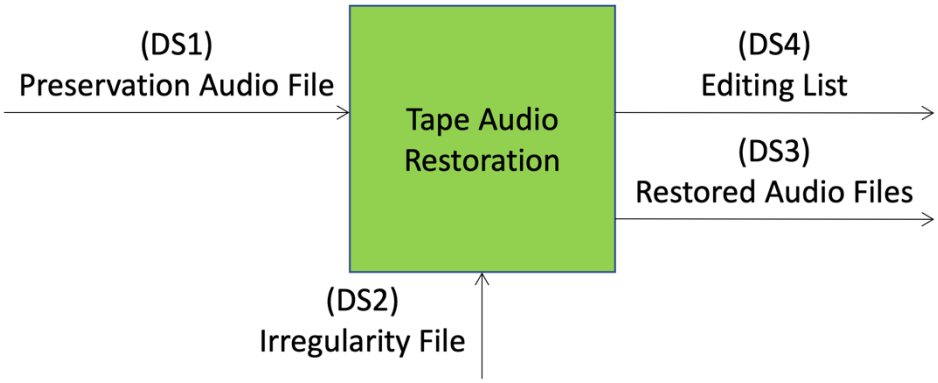

Figure 13 – *Tape Audio Restoration.*

After the Tests, Conformance Tester shall fill out Table 33.

Table 33 – Conformance Testing form of Audio Recording Preservation (ARP) Tape Audio Restoration.

<table>
<tr><td>Conformance Tester ID</td><td>Unique Conformance Tester Identifier assigned by MPAI</td></tr>
<tr><td>Standard, Use Case ID and Version</td><td>Standard ID and Use Case ID, Version and Profile of the standard in the form “CAE:ARP:1:0”.</td></tr>
<tr><td>Name of AIM</td><td>Tape Audio Restoration</td></tr>
<tr><td>Implementer ID</td><td>Unique Implementer Identifier assigned by MPAI Store.</td></tr>
<tr><td>AIM Implementation Version</td><td>Unique Implementation Identifier assigned by Implementer.</td></tr>
<tr><td>Neural Network Version*</td><td>Unique Neural Network Identifier assigned by Implementer.</td></tr>
<tr><td>Identifier of Conformance Testing Dataset</td><td>Unique Dataset Identifier assigned by MPAI Store.</td></tr>
<tr><td>Test ID</td><td>Unique Test Identifier assigned by Conformance Tester.</td></tr>
<tr><td>Actual output</td><td>Actual output provided as a matrix of n rows containing output assertions.

For example:
<table>
<tr><th>#</th><th>Irr</th><th>Editing List errors</th><th>Duration</th><th>Xcorr</th><th>RMSE</th><th>PSD</th><th>Final assertion</th></tr>
<tr><td>1</td><td>SB, ESV, SSV</td><td>0, 1, 2…</td><td>T/F</td><td>T/F</td><td>T/F</td><td>T/F</td><td>T/F</td></tr>
<tr><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td></tr>
<tr><td>n</td><td>SB, SSV</td><td>0, 1, 2…</td><td>T/F</td><td>T/F</td><td>-</td><td>-</td><td>T/F</td></tr>
</table>
Final evaluation: T/F

Legend:
- #: CT dataset tuple number.
- Irr: Irregularity types present on the Preservation Audio File
- Editing List errors: number of edits incorrectly performed. It has negative impact if different from 0
- Duration: flag to check the Restored Audio Files duration
- Xcorr: flag to check the cross-correlation measures
- RMSE: flag to check the RMSE measures (only for ESV)
- PSD: flag to check the PSD measures (only for ESV)
- Final assertion: AND operation between previous results</td></tr>
<tr><td>Execution time*</td><td>Duration of test execution.</td></tr>
<tr><td>Test comment*</td><td>Comments on test results and possible needed actions.</td></tr>
<tr><td>Test Date</td><td>yyyy/mm/dd.</td></tr>
</table>

** Optional field*

5.3.6 Audio Recording Preservation (ARP) Packager

Table 34 gives the input/output data of the Packager.

Table 34 – I/O Data of Audio Recording Preservation (ARP) Packager.

AIM	Input Data	Output Data
Pack-ager	Preservation Audio File Preservation Audio-Visual File Restored Audio Files Editing List Irregularity File Irregularity Images	Access Copy Files Preservation Master Files

Table 35 gives the Audio Recording Preservation (ARP) *Packager* Means and how they are used.

Table 35 – AIM Means and use of Audio Recording Preservation (ARP) Packager.

Means	Actions
Conform-ance Test-ing Dataset	DS1: *n* Preservation Audio Files. DS2: *n* Preservation Audio-Visual Files related to DS1. DS3: *n* Restored Audio Files arrays related to DS1 coming from Tape Audio Restoration. DS4: *n* Editing Lists related to DS3 coming from Tape Audio Restoration. DS5: *n* Irregularity Files related to DS1 coming from Tape Irregularity Classifier. DS6: *n* Irregularity Images related to DS5 coming from Tape Irregularity Classifier. DS7: *n* Access Copy Files. DS8: *n* Preservation Master Files.
Procedure	1. Feed Packager under test with DS1, DS2, DS3, DS4, DS5 and DS6. 2. Compare the output Access Copy Files with DS7. 3. Compare the output Preservation Master Files with DS8.
Evaluation	For a given input tuple, verify that: 1. The output Access Copy Files contain the Restored Audio Files, the Editing List, the Irregularity File and the set of Irregularity Images in a .zip file and is therefore equal to DS7. 2. The output Preservation Master Files contain the Preservation Audio File, the Preservation Audio-Visual File with the audio of the Preservation Audio File, the Irregularity File, and the Irregularity Images, and is therefore equal to DS8. An error on any of the output arrays will make the Packager under test not conformant.

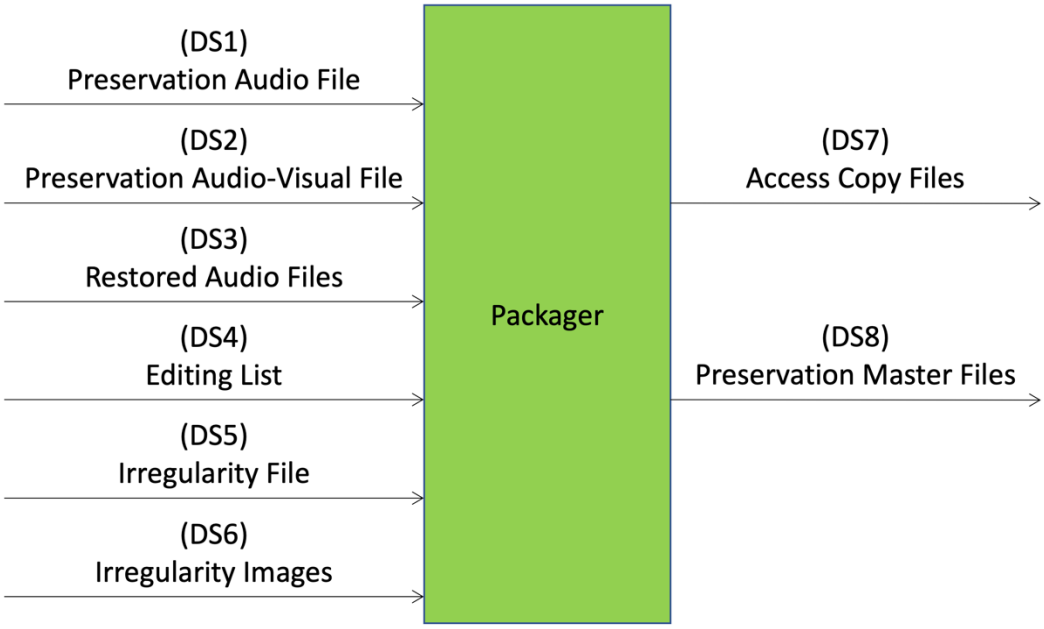

Figure 14 – *Packager.*

After the Tests, Conformance Tester shall fill out Table 36.

Table 36 – Conformance Testing form of Audio Recording Preservation (ARP) Packager.

<table>
<tr><td>Conformance Tester ID</td><td>Unique Conformance Tester Identifier assigned by MPAI</td></tr>
<tr><td>Standard, Use Case ID and Version</td><td>Standard ID and Use Case ID, Version and Profile of the standard in the form “CAE:ARP:1:0”.</td></tr>
<tr><td>Name of AIM</td><td>Packager</td></tr>
<tr><td>Implementer ID</td><td>Unique Implementer Identifier assigned by MPAI Store.</td></tr>
<tr><td>AIM Implementation Version</td><td>Unique Implementation Identifier assigned by Implementer.</td></tr>
<tr><td>Neural Network Version*</td><td>Unique Neural Network Identifier assigned by Implementer.</td></tr>
<tr><td>Identifier of Conformance Testing Dataset</td><td>Unique Dataset Identifier assigned by MPAI Store.</td></tr>
<tr><td>Test ID</td><td>Unique Test Identifier assigned by Conformance Tester.</td></tr>
<tr><td>Actual output</td><td>Actual output provided as a matrix of n rows containing output assertions.
<table>
<tr><th>Output</th><th>Files</th><th>1</th><th>…</th><th>n</th></tr>
<tr><td rowspan="4">Access Copy Files</td><td>Restored Audio Files</td><td>T/F</td><td>…</td><td>T/F</td></tr>
<tr><td>Editing List</td><td>T/F</td><td>…</td><td>T/F</td></tr>
<tr><td>Irregularity File</td><td>T/F</td><td>…</td><td>T/F</td></tr>
<tr><td>Irregularity Images</td><td>T/F</td><td>…</td><td>T/F</td></tr>
<tr><td>Preservation Master Files</td><td>Preservation Audio File</td><td>T/F</td><td>…</td><td>T/F</td></tr>
</table></td></tr>
</table>

	Preservation Audio-Visual File	T/F	…	T/F
	Irregularity File	T/F	…	T/F
	Irregularity Images	T/F	…	T/F
	Final assertion: T/F			
Execution time*	Duration of test execution.			
Test comment*	Comments on test results and possible needed actions.			
Test Date	yyyy/mm/dd.			

** Optional field*

5.4 Speech Restoration System (CAE-SRS)

The normative Architecture of the Speech Restoration System (CAE-SRS) is given by *Figure 15*.

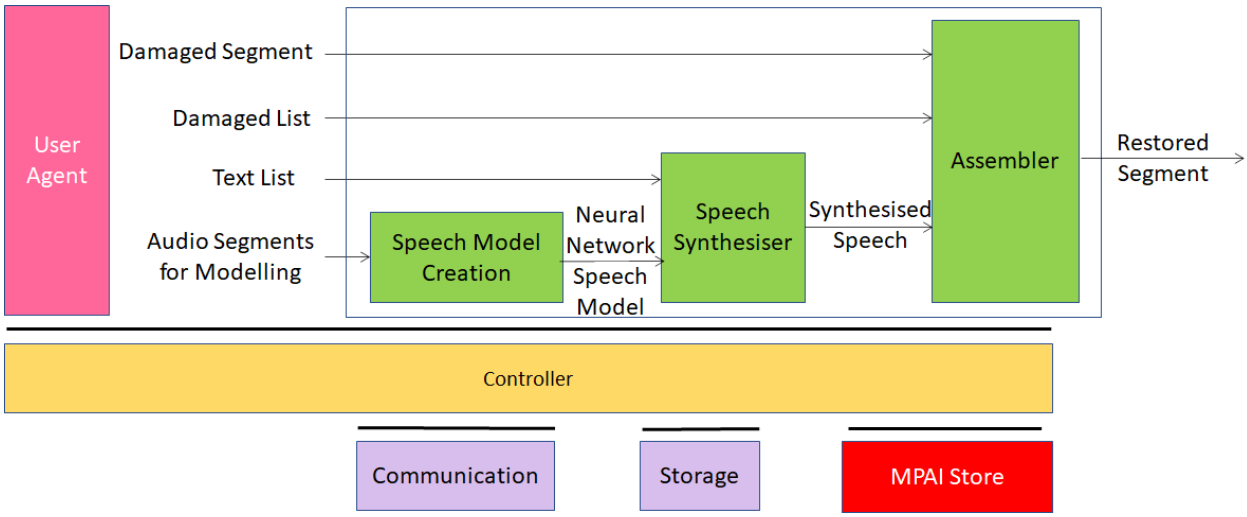

Figure 15 – Reference Model of Speech Restoration System (CAE-SRS)

The input/output data format are specified in 5.1.2 [3].

5.4.1 Speech Restoration System (SRS) AIMs

*Table 37**Error! Reference source not found.*** gives the AIMs and the input/output data of Speech Restoration System (SRS).

Table 37 – AIMs and their I/O data of Speech Restoration System (SRS)

AIM	Input Data	Output Data
Speech Model Creation	Audio Segments for Modelling	Neural Network Speech Model
Speech Synthesiser	Text List Neural Network Speech Model	Synthesised Speech
Assembler	Damaged Segments Damaged List Synthesised Speech	Restored Segment

5.4.2 Speech Model Creation

Table 38 gives the input/output data of Speech Model Creation.

Table 38 – I/O Data of Speech Restoration System (CAE-SRS) Speech Model Creation

AIM	Input Data	Output Data
Speech Model Creation	Audio Segments for Modelling	Neural Network Speech Model

Table 39 gives the Speech Restoration System *(CAE-SRS) Speech Model Creation* Means and how they are used.

Table 39 - AIM Means and use of Speech Restoration System (CAE-SRS) Speech Model Creation

Means	Actions
Conformance Testing Dataset	DS1: a set of *n* Audio Segments, suitable input for creation of a Neural Network Speech Model for a specific speaker.
Procedure	1. Pass Audio Segments for Modelling to Speech Module Creation AIM as input, according to its declared standard procedure. 2. Provide resulting Neural Network Speech Model as input to the reference Speech Synthesiser AIM (ID: *ss*). 3. Synthesise all texts in canonical Text List.
Evaluation	1. Evaluate synthesis quality using Perception Evaluation of Speech Quality (PESQ). 2. If the score is above a threshold of 2.0, the Speech Model Creation AIM is judged adequate. 3. If the quality is below threshold, the submitter of Speech Model Creation is given the opportunity to submit an implementation of Speech Synthesiser. 4. The MPAI Store will test the combination of the two submitted AIMs. 5. If the quality of the output of the submitted combination is above threshold, Speech Model Creation passes the Conformance Test as long as the corresponding Speech Synthesiser are made available to the MPAI Store. 6. Else, Speech Model Creation doesn't pass the Conformance Test.

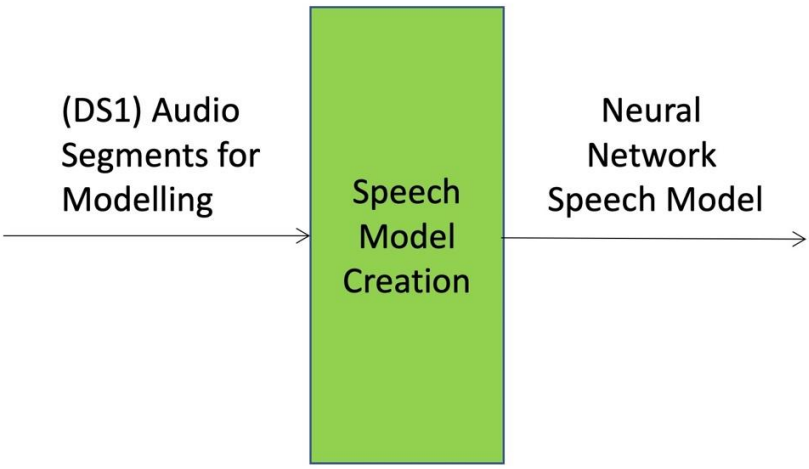

Figure 16 - Speech Model Creation.

After the Tests, Conformance Tester shall fill out *Table 40.*

Table 40 - Conformance Testing form of Speech Restoration System (SRS) Speech Model Creation.

<table>
<tr><td>Conformance Tester ID</td><td>Unique Conformance Tester Identifier assigned by MPAI</td></tr>
<tr><td>Standard, Use Case ID and Version</td><td>Standard ID and Use Case ID, Version and Profile of the standard in the form “CAE:SRS:1:0”.</td></tr>
<tr><td>Name of AIM</td><td>Speech Model Creation</td></tr>
<tr><td>Implementer ID</td><td>Unique Implementer Identifier assigned by MPAI Store.</td></tr>
<tr><td>AIM Implementation Version</td><td>Unique Implementation Identifier assigned by Implementer.</td></tr>
<tr><td>Neural Network Version*</td><td>Unique Neural Network Identifier assigned by Implementer.</td></tr>
<tr><td>Identifier of Conformance Testing Dataset</td><td>Unique Dataset Identifier assigned by MPAI Store.</td></tr>
<tr><td>Test ID</td><td>Unique Test Identifier assigned by Conformance Tester.</td></tr>
<tr><td>Actual output</td><td>Actual output provided as a matrix of n rows containing output assertions.

For example:

<table>
<tr><th>#</th><th>Input</th><th>Final assertion</th></tr>
<tr><td>1</td><td>DS1[1]</td><td>T/F</td></tr>
<tr><td>…</td><td>…</td><td>…</td></tr>
<tr><td>n</td><td>DS1[n]</td><td>T/F</td></tr>
</table>

Final evaluation: T/F

Legend:
- #: Speech Model Creation input dataset tuple number.
- DS1: Audio Segments for Modelling
- Final assertion: T if Neural Network Speech Model is well-formed, else F
- Final evaluation: T if all Final assertions are T, else F</td></tr>
<tr><td>Execution time*</td><td>Duration of test execution.</td></tr>
<tr><td>Test comment*</td><td>Comments on test results and possible needed actions.</td></tr>
<tr><td>Test Date</td><td>yyyy/mm/dd.</td></tr>
</table>

** Optional field*

5.4.3 Speech Synthesiser

Table 41 gives the input/output data of Speech Synthesiser.

Table 41 - I/O Data of Speech Restoration System (CAE-SRS) Speech Synthesiser

AIM	Input Data	Output Data
Speech Synthesiser	Text List Neural Network Speech Model	Synthesised Speech

Table 42 gives the *Speech Restoration System (CAE-SRS) Speech Synthesiser* Means and how they are used.

Table 42 - AIM Means and use of Speech Restoration System (CAE-SRS) Speech Synthesiser

Means	Actions
Conformance Testing Dataset	DS1: a dataset of *n* canonical Text Lists. DS2: a dataset of *n* canonical Neural Network Speech Models.
Procedure	1. Pass canonical Text Lists and canonical Neural Network Speech Models to Speech Synthesiser AIM as input, according to its declared standard procedure. 2. Synthesise all texts in canonical Text List.
Evaluation	1. Evaluate synthesis quality using Perception Evaluation of Speech Quality (PESQ). 2. If score passes declared threshold, Speech Synthesiser is judged adequate. 3. Else, Speech Synthesiser doesn't pass the Conformance Test.

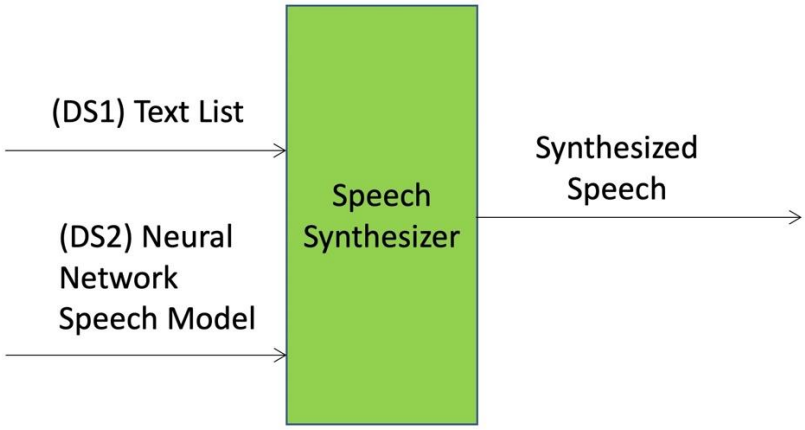

Figure 17 - Speech Synthesiser.

After the Tests, Conformance Tester shall fill out *Table 43.*

Table 43 - Conformance Testing form of Speech Restoration System (SRS) Speech Synthesiser.

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:SRS:1:0".
Name of AIM	Speech Synthesiser
Implementer ID	Unique Implementer Identifier assigned by MPAI Store.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Conformance Testing Dataset	Unique Dataset Identifier assigned by MPAI Store.
Test ID	Unique Test Identifier assigned by Conformance Tester.
Actual output	Actual output provided as a matrix of *n* rows containing output assertions. For example:

	<table><tr><th>#</th><th>Input</th><th>Final assertion</th></tr><tr><td>1</td><td>DS1[1], DS2[1]</td><td>T/F</td></tr><tr><td>…</td><td>…</td><td>…</td></tr><tr><td>*n*</td><td>DS1[*n*], DS2[*n*]</td><td>T/F</td></tr></table> Final evaluation: T/F Legend: - **#**: Speech Synthesiser input dataset tuple number. - **DS1**: Text List - **DS2**: Neural Network Speech Model - **Final assertion**: T if Synthesised Speech is audible and intelligible, else F - **Final evaluation**: T if all **Final assertions** are T, else F
Execution time*	Duration of test execution.
Test comment*	Comments on test results and possible needed actions.
Test Date	yyyy/mm/dd.

** Optional field*

5.4.4 Assembler

Table 44 gives the input/output data of Assembler.

Table 44 - I/O Data of Speech Restoration System (CAE-SRS) Assembler

AIM	Input Data	Output Data
Assembler	Damaged Segment Damaged List Synthesised Speech	Restored Segment

Table 45 gives the *Speech Restoration System (CAE-SRS) Assembler* Means and how they are used.

Table 45 - AIM Means and use of Speech Restoration System (CAE-SRS) Assembler

Means	Actions
Conformance Testing Dataset	DS1: a canonical set of *n* Damaged Segments DS2: a canonical set of *n* Damaged Lists DS3: a canonical set of *n* Synthesised Speeches.
Procedure	1. Pass DS1, DS2 and DS3 to Assembler, according to its declared standard Procedure. 2. Perform all specified assembly operations: Synthesised Speech results shall replace all bad sections of Damaged Segment as specified by Damaged List.
Evaluation	1. Restored Segment shall be evaluated for quality using Perception Evaluation of Speech Quality (PESQ). Restoration shall be seamless, so that listeners are unable to reliably identify locations of repaired sections. 2. If the scores exceed a declared threshold, Assembler is judged adequate. 3. Else, Assembler doesn't pass the Conformance Test.

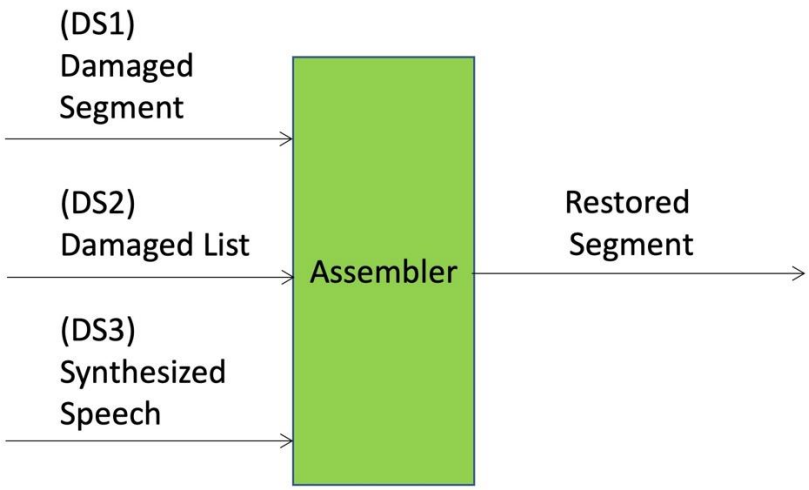

Figure 18 - Assembler.

After the Tests, Conformance Tester shall fill out *Table 46.*

Table 46 - Conformance Testing form of Speech Restoration System (SRS) Assembler.

<table>
<tr><td>Conformance Tester ID</td><td>Unique Conformance Tester Identifier assigned by MPAI</td></tr>
<tr><td>Standard, Use Case ID and Version</td><td>Standard ID and Use Case ID, Version and Profile of the standard in the form “CAE:SRS:V:P”.</td></tr>
<tr><td>Name of AIM</td><td>Assembler</td></tr>
<tr><td>Implementer ID</td><td>Unique Implementer Identifier assigned by MPAI Store.</td></tr>
<tr><td>AIM Implementation Version</td><td>Unique Implementation Identifier assigned by Implementer.</td></tr>
<tr><td>Neural Network Version*</td><td>Unique Neural Network Identifier assigned by Implementer.</td></tr>
<tr><td>Identifier of Conformance Testing Dataset</td><td>Unique Dataset Identifier assigned by MPAI Store.</td></tr>
<tr><td>Test ID</td><td>Unique Test Identifier assigned by Conformance Tester.</td></tr>
<tr><td>Actual output</td><td>Actual output provided as a matrix of n rows containing output assertions.

For example:

<table>
<tr><td>#</td><td>Input tuple</td><td>Final assertion</td></tr>
<tr><td>1</td><td>DS1[1],
DS2[1],
DS3[1]</td><td>T/F</td></tr>
<tr><td>…</td><td>…</td><td>…</td></tr>
<tr><td>n</td><td>DS1[n],
DS2[n],
DS3[n]</td><td>T/F</td></tr>
</table>

Final evaluation: T/F

Legend:
- #: Assembler input dataset tuple number.
- DS1: Damaged Segment (within which damaged sub-segments may be listed)</td></tr>
</table>

	- **DS2**: Damaged List (of damaged sub-segments within current Damaged Segment) - **DS3**: Synthesised Speech (list of synthesised sub-segments corresponding to damaged sub-segments of DL) - **Final assertion**: T if Restored Segment is well-formed (single audio file without audible interruptions or gaps, produced without error messages or breaks), else F - **Final evaluation**: T if all **Final assertions** are T, else F
Execution time*	Duration of test execution.
Test comment*	Comments on test results and possible needed actions.
Test Date	yyyy/mm/dd.

** Optional field*

5.5 Enhanced Audioconference Experience (CAE-EAE)

The normative Architecture of the Enhanced Audioconference Experience (CAE-EAE) is given by *Figure 19*.

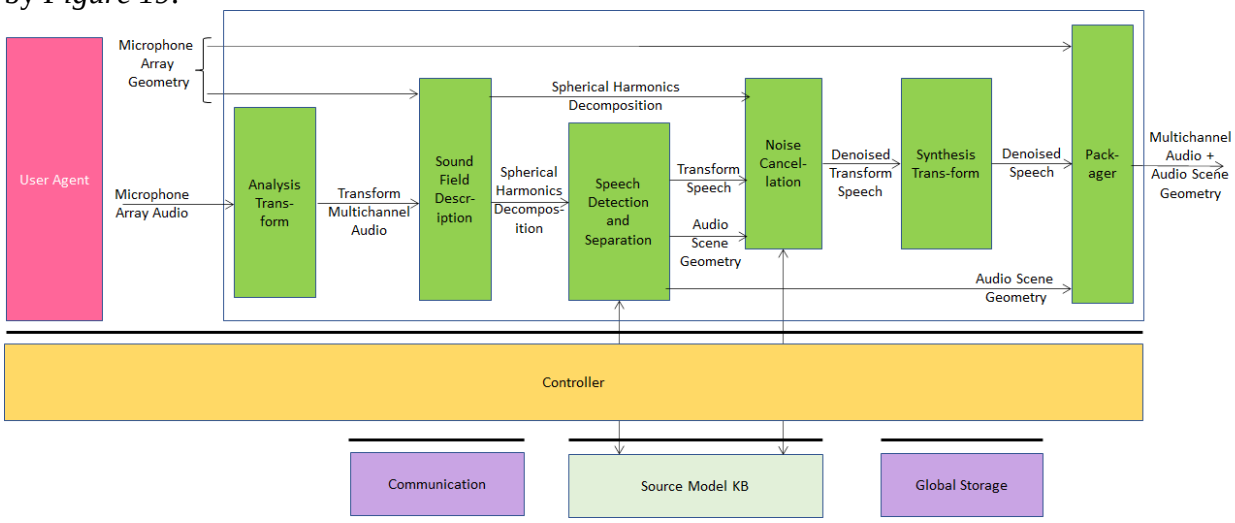

Figure 19 – Reference Model of Enhanced Audioconference Experience (CAE-EAE)

The input/output data format are specified in 5.4 [3].

5.5.1 Enhanced Audioconference Experience (CAE-EAE) AIMs

Table 47 gives the AIMs and the input/output data of Enhanced Audioconference Experience (CAE-EAE).

Table 47 – AIMs and their I/O data of Enhanced Audioconference Experience (CAE-EAE)

AIM	Input Data	Output Data
Analysis Transform	Microphone Array Audio	Transform Multichannel Audio
Sound Field Description	Transform Multichannel Audio Microphone Array Geometry	Spherical Harmonics Decomposition (SHD)
Speech Detection and Separation	SHD	Transform Speech Audio Scene Geometry

Noise Cancellation	SHD Transform Speech Audio Scene Geometry	Denoised Transform Speech
Synthesis Transform	Denoised Transform Speech	Denoised Speech
Packager	Denoised Speech Audio Scene Geometry Microphone Array Geometry	Multichannel Audio + Audio Scene Geometry

5.5.2 Enhanced Audioconference Experience (CAE-EAE) Analysis Transform

Table 48 gives the input/output data of the Analysis Transform.

Table 48 – I/O Data of Enhanced Audioconference Experience (CAE-EAE) Analysis Transform

AIM	Input Data	Output Data
Analysis Transform	Microphone Array Audio	Transform Multichannel Audio

Table 49 gives the *Enhanced Audioconference Experience (CAE-EAE) Analysis Transform* Means and how they are used.

Table 49 – AIM Means and use of Enhanced Audioconference Experience (CAE-EAE) Analysis Transform

Means	Actions
Conformance Testing Dataset	DS1: *n* Test files including Microphone Array Audio in Interleaved Multichannel Audio format. DS2: *n* Expected Output files including data in Transform Interleaved Multichannel Audio format.
Procedure	1. Feed the AIM under test with the Test files (DS1). 2. Analyse the Transform Multichannel Audio with the Expected Output files (DS2).
Evaluation	1. Check the Transform Multichannel Audio data format with the given Expected Output files format. 2. Calculate the peak-to-peak Amplitude (A) of each Audio block in the Expected Output files. 3. Calculate the RMSE of each Audio block by comparing the Transform Multichannel Audio (x) with the Expected Output files (y). 4. Accept the AIM under test if, for each audio block, these two conditions are satisfied: a. Data format of the Transform Multichannel Audio is the same with the Expected Output Files format and b. RMSE < A* 0.1%.

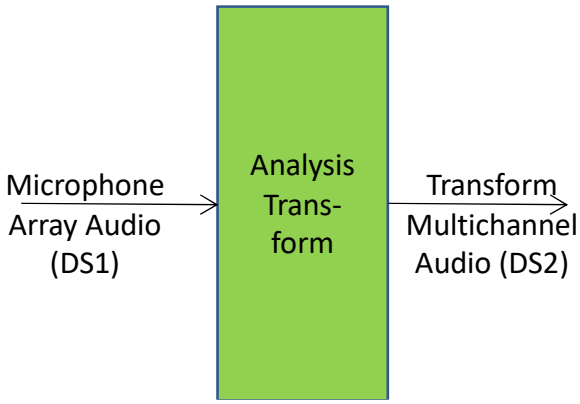

Figure 20 – Analysis Transform Testing Flow

After the Tests, Conformance Tester shall fill out *Table 50*.

Table 50 – Conformance Testing form of Enhanced Audioconference Experience (CAE-EAE) Analysis Transform

<table>
<tr><td>Conformance Tester ID</td><td>Unique Conformance Tester Identifier assigned by MPAI</td></tr>
<tr><td>Standard, Use Case ID and Version</td><td>Standard ID and Use Case ID, Version and Profile of the standard in the form “CAE:EAE:1:0”.</td></tr>
<tr><td>Name of AIM</td><td>Analysis Transform</td></tr>
<tr><td>Implementer ID</td><td>Unique Implementer Identifier assigned by MPAI Store.</td></tr>
<tr><td>AIM Implementation Version</td><td>Unique Implementation Identifier assigned by Implementer.</td></tr>
<tr><td>Neural Network Version*</td><td>Unique Neural Network Identifier assigned by Implementer.</td></tr>
<tr><td>Identifier of Conformance Testing Dataset</td><td>Unique Dataset Identifier assigned by MPAI Store.</td></tr>
<tr><td>Test ID</td><td>Unique Test Identifier assigned by Conformance Tester.</td></tr>
<tr><td>Actual output</td><td>The Conformance Tester will provide the following matrix containing a limited number of input records (n) with the corresponding outputs. If an input record fails, the tester would specify the reason why the test case fails.
<table>
<tr><th>Input data (DS1)</th><th>Expected Output Data (DS2)</th><th>Data Format</th><th>RMSE</th></tr>
<tr><td>Microphone Array Audio ID_1</td><td>Transform Multichannel Audio ID_1</td><td>T/F</td><td>< A*0.1%</td></tr>
<tr><td>Microphone Array Audio ID_2</td><td>Transform Multichannel Audio ID_2</td><td>T/F</td><td>< A*0.1%</td></tr>
<tr><td>Microphone Array Audio ID_3</td><td>Transform Multichannel Audio ID3</td><td>T/F</td><td>< A*0.1%</td></tr>
<tr><td>…</td><td>…</td><td>…</td><td>…</td></tr>
<tr><td>Microphone Array Audio ID_n</td><td>Transform Multichannel Audio ID_n</td><td>T/F</td><td>< A*0.1%</td></tr>
</table></td></tr>
</table>

	Final evaluation: T/F Denoting with i, $0 \leq i < n = 50$, the record number in DS1 and DS2, the matrices reflect the results obtained with input records i with the corresponding outputs i. DS1 | DS2 | Analysis Transform output value (obtained through the AIM under test) DS1[i] | DS2[i] | AnalysisTransform[i]
Execution time*	Duration of test execution.
Test comment*	Comments on test results and possible needed actions.
Test Date	yyyy/mm/dd.

** Optional field*

5.5.3 Enhanced Audioconference Experience (CAE-EAE) Sound Field Description

Table 51 Table 51 gives the input/output data of the Sound Field Description.

Table 51 – I/O Data of Enhanced Audioconference Experience (CAE-EAE) Sound Field Description

AIM	Input Data	Output Data
Sound Field Description	Transform Multichannel Audio Microphone Array Geometry	SHD

Table 52 Table 52 gives the *Enhanced Audioconference Experience (CAE-EAE) Sound Field Description* Means and how they are used.

Table 52 – AIM Means and use of Enhanced Audioconference Experience (CAE-EAE) Sound Field Description

Means	Actions
Conformance Testing Dataset	DS1: *n* Test files containing real recordings or simulations structured in Transform Multichannel Audio format. DS2: *n* Microphone Array Geometry associated with the real recordings or simulations. DS3: *n* Expected Output files including data in SHD format.
Procedure	1. Feed the AIM under test with the Test files (DS1) and their associated Microphone Array Geometry (DS2). 2. Analyse the SHD with the Expected Output files (DS3).
Evaluation	1. Check the output SHD data format with the given Expected Output files format. 2. Calculate the peak-to-peak Amplitude (A) value of each Audio block in the Expected Output files.

	3. Calculate the RMSE of each Audio block in SHD by comparing the output (*x*) with the Expected Output files (*y*). 4. Accept the AIM under test if, for each audio block, these two conditions are satisfied: a. Data format of the SHD is the same with the Expected Output Files and b. RMSE < A * 0.1%

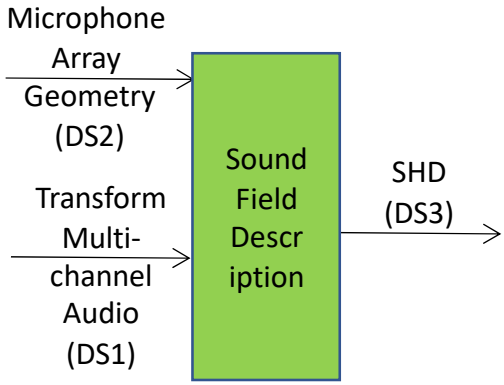

Figure 21 - Sound Field Description Testing Flow

After the Tests, Conformance Tester shall fill out *Table 53*.

Table 53 – Conformance Testing form of Enhanced Audioconference Experience (CAE-EAE) Sound Field Description

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:EAE:1:0".
Name of AIM	Sound Field Description
Implementer ID	Unique Implementer Identifier assigned by MPAI Store.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Conformance Testing Dataset	Unique Dataset Identifier assigned by MPAI Store.
Test ID	Unique Test Identifier assigned by Conformance Tester.

Actual output	The Conformance Tester will provide the following matrix containing a limited number of input records (n) with the corresponding outputs. If an input record fails, the tester would specify the reason why the test case fails.

Input data (DS1, DS2)	Expected Output Data (DS3)	Data Format	RMSE
Transform Multichannel Audio ID_1 Microphone Array Geom-etry ID_1	SHD ID_1	T/F	< A * 0.1%
Transform Multichannel Audio ID_2 Microphone Array Geom-etry ID_2	SHD ID_2	T/F	< A * 0.1%
Transform Multichannel Audio ID_3 Microphone Array Geom-etry ID_3	SHD ID_3	T/F	< A * 0.1%
…	…	…	…
Transform Multichannel Audio ID_n Microphone Array Geom-etry ID_n	SHD ID_n	T/F	< A * 0.1%

Final evaluation: T/F

Denoting with i, $1 \leq i < n = 50$, the record number in DS1, DS2, and DS3, the matrices reflect the results obtained with input records i with the corresponding outputs i.

DS1	DS2	DS3	Sound Field Description output value (obtained through the AIM under test)
DS1[i]	DS2[i]	DS3[i]	SoundFieldDescription[i]

Execution time*	Duration of test execution.
Test comment*	Comments on test results and possible needed actions.
Test Date	yyyy/mm/dd.

** Optional field*

5.5.4 Enhanced Audioconference Experience (CAE-EAE) Speech Detection and Separation

Table 54 gives the input/output data of the Speech Detection and Separation.

Table 54- I/O Data of Enhanced Audioconference Experience (CAE-EAE) Speech Detection and Separation

AIM	Input Data	Output Data
Speech Detection and Separation	SHD	Transform Speech Audio Scene Geometry

Table 55 gives the *Enhanced Audioconference Experience (CAE-EAE) Speech Detection and Separation* Means and how they are used.

Table 55 – AIM Means and use of Enhanced Audioconference Experience (CAE-EAE) Speech Detection and Separation

Means	Actions
Conformance Testing Dataset	The Conformance Testing Dataset is composed b: DS1: *n* Test files containing SHD. DS2: *n* Expected Transform Speech Files. DS3: *n* Expected Audio Scene Geometry.
Procedure	1. Feed the AIM under test with the Test files. 2. Analyse the Audio Scene Geometry. 3. Analyse Transform Speech Files.
Evaluation	1. Control the Audio Scene Geometry with the Expected Audio Scene Geometry: a. Count the number of objects in the Audio Scene Geometry. b. Calculate the angle difference (AD) in degrees between the objects (u) in the Audio Scene Geometry and the objects (v) in the Expected Audio Scene Geometry. 2. Compare the number of Audio Blocks in the Expected Transform Speech with the number of Audio Blocks in the Transform Speech Files. 3. Calculate Signal to Interference Ratio (SIR), Signal to Distortion Ratio (SDR), and Signal to Artefacts Ratio (SAR) between the Expected and Output Transform Speech Files [12]. 4. Accept the AIM under test if these four conditions are satisfied: a. The number of speech objects in the Audio Scene Geometry is equal to the number of speech objects in the Expected Audio Scene Geometry. b. The number of Audio Blocks in the Transform Speech is equal to the number of Audio Blocks in the Expected Transform Speech. c. Compare each Speech Object in the Audio Scene Geometry with the Speech Object in the Expected Audio Scene Geometry. i. Each object's AD between the Expected and Output is less than 5 degrees. d. Compare each Speech Object in the Transform Speech with the Speech Object in the Expected Transform Speech.

	i. If the room reverb time (T60) is greater than 0.5 seconds. 1. Each object's SIR between the Expected and Output is greater than or equal to 10 dB. 2. Each object's SDR between the Expected and Output is greater than or equal to 3 dB. 3. Each object's SAR between the Expected and Output is greater than or equal to 3 dB. ii. If the room reverb time (T60) is less than 0.5 seconds. 1. Each object's SIR between the Expected and Output is greater than or equal to 15 dB. 2. Each object's SDR between the Expected and Output is greater than or equal to 6 dB. 3. Each object's SAR between the Expected and Output is greater than or equal to 6 dB.

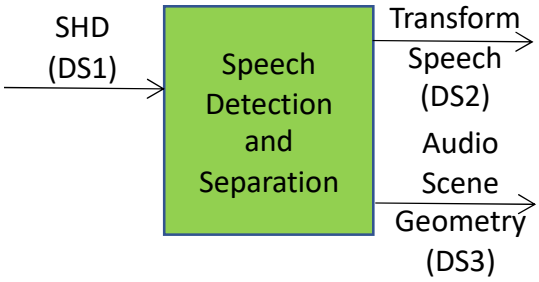

Figure 22 - Speech Detection and Separation Testing Flow

After the Tests, Conformance Tester shall fill out *Table 56.Table 56*

Table 56 – Conformance Testing form of Enhanced Audioconference Experience (CAE-EAE) Speech Detection and Separation

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:EAE:1:0".
Name of AIM	Speech Detection and Separation
Implementer ID	Unique Implementer Identifier assigned by MPAI Store.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Conformance Testing Dataset	Unique Dataset Identifier assigned by MPAI Store.
Test ID	Unique Test Identifier assigned by Conformance Tester.
Actual output	The Conformance Tester will provide the following matrix containing a limited number of input records (*n*) with the corresponding outputs. If an input record fails, the tester would specify the reason why the test case fails.

Input data	Expected Output Data	Data Format	Audio Scene	Source Separation

(DS1)	(DS2, DS3)		Geometry	Metrics
SHD ID_1	Transform Speech ID_1 Audio Scene Geometry ID_1	T/F	T/F	T/F
SHD ID_2	Transform Speech ID_2 Audio Scene Geometry ID_2	T/F	T/F	T/F
SHD ID_3	Transform Speech ID_3 Audio Scene Geometry ID_3	T/F	T/F	T/F
…	…	…	…	…
SHD ID_n	Transform Speech ID_n Audio Scene Geometry ID_n	T/F	T/F	T/F

Final evaluation : T/F

Denoting with i, $1 \leq i < n = 50$, the record number in DS1, DS2, and DS3, the matrices reflect the results obtained with input records i with the corresponding outputs i.

DS1	DS2	DS3	Sound Field Description output value (obtained through the AIM under test)
DS1[i]	DS2[i]	DS3[i]	SpeechDetectionandSeparation[i]

Execution time*	Duration of test execution.
Test comment*	Comments on test results and possible needed actions.
Test Date	yyyy/mm/dd.

** Optional field*

5.5.5 Enhanced Audioconference Experience (CAE-EAE) Noise Cancellation

Table 57 gives the input/output data of the Noise Cancellation.

Table 57 – I/O Data of Enhanced Audioconference Experience (CAE-EAE) Noise Cancellation

AIM	Input Data	Output Data
Noise Cancellation	Transform Speech Audio Scene Geometry	Denoised Transform Speech

Table 58 gives the *Enhanced Audioconference Experience (CAE-EAE) Noise Cancellation* Means and how they are used.

Table 58 – AIM Means and use of Enhanced Audioconference Experience (CAE-EAE) Noise Cancellation

Means	**Actions**
Conformance Testing Dataset	DS1: *n* Test files containing SHD. DS2: *n* Test files containing Transform Speech. DS3: *n* Test files containing Audio Scene Geometry. DS4: *n* Expected Denoised Transform Speech.
Procedure	1. Feed the AIM under test with the Test files (DS1, DS2, DS3). 2. Analyse the Denoised Transform Speech (DS4).
Evaluation	1. Compare the number of Audio Blocks in the Expected Denoised Transform Speech with the number of Audio Blocks in the Denoised Transform Speech Files. 2. Compute Perception Evaluation of Speech Quality (PESQ) between the Expected and Output Denoised Transform Speech Files [6]. 3. Accept the AIM under test if these two conditions are satisfied: a. The number of Audio Blocks in the Denoised Transform Speech is the same with the number of Audio Blocks in the Expected Denoised Transform Speech. b. Compare each Denoised Transform Speech with the Expected Denoised Transform Speech. c. If the room reverb time (T60) is greater than 0.5 seconds. i. Each object's PESQ between the Expected and Output is greater than P=2.0. d. If the room reverb time (T60) is smaller than 0.5 seconds. i. Each object's PESQ between the Expected and Output is greater than P=3.0.

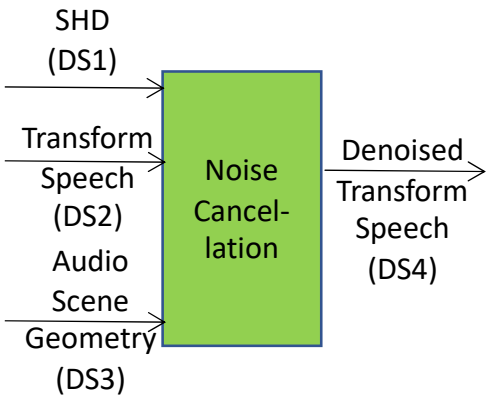

Figure 23 - Noise Cancellation Testing Flow

After the Tests, Conformance Tester shall fill out *Table 59.Table 59*

Table 59 – Conformance Testing form of Enhanced Audioconference Experience (CAE-EAE) Noise Cancellation

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:EAE:1:0".
Name of AIM	Noise Cancellation
Implementer ID	Unique Implementer Identifier assigned by MPAI Store.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Conformance Testing Dataset	Unique Dataset Identifier assigned by MPAI Store.
Test ID	Unique Test Identifier assigned by Conformance Tester.
Actual output	The Conformance Tester will provide the following matrix containing a limited number of input records (*n*) with the corresponding outputs. If an input record fails, the tester would specify the reason why the test case fails.

Input data (DS1, DS2, DS3)	Expected Output Data (DS4)	Data Format	PESQ Score
SHD ID_1 Transform Speech ID_1 Audio Scene Geometry ID_1	Denoised Transform Speech ID_1	T/F	> P
SHD ID_2 Transform Speech ID_2 Audio Scene Geometry ID_2	Denoised Transform Speech ID_2	T/F	> P
SHD ID_3 Transform Speech ID_3 Audio Scene Geometry ID_3	Denoised Transform Speech ID_3	T/F	> P
…	…	…	…
SHD ID_n Transform Speech ID_n Audio Scene Geometry ID_n	Denoised Transform Speech ID_n	T/F	> P

	Final evaluation : T/F Denoting with i, $1 \leq i < n = 50$, the record number in DS1, DS2, and DS3, the matrices reflect the results obtained with input records i with the corresponding outputs i. DS1 | DS2 | DS3 | DS4 | Noise Cancellation output value (obtained through the AIM under test) DS1[i] | DS2[i] | DS3[i] | DS4[i] | NoiseCancellation[i]
Execution time*	Duration of test execution.
Test comment*	Comments on test results and possible needed actions.
Test Date	yyyy/mm/dd.

** Optional field*

5.5.6 Enhanced Audioconference Experience (CAE-EAE) Synthesis Transform

Table 60 gives the input/output data of the Synthesis Transform.

Table 60 – I/O Data of Enhanced Audioconference Experience (CAE-EAE) Synthesis Transform

AIM	Input Data	Output Data
Synthesis Transform	Denoised Transform Speech	Denoised Speech

Table 61 gives the *Enhanced Audioconference Experience (CAE-EAE) Synthesis Transform* Means and how they are used.

Table 61 – AIM Means and use of Enhanced Audioconference Experience (CAE-EAE) Synthesis Transform

Means	Actions
Conformance Testing Dataset	DS1: *n* Test files including data in Denoised Transform Speech format. DS2: *n* Expected Output files including data in Denoised Speech format.
Procedure	1. Feed the AIM under test with the Test files (DS1). 2. Analyse the Denoised Speech with the Expected Output files (DS2).
Evaluation	1. Check the Denoised Speech data format with the given Expected Output files format. 2. Calculate the peak-to-peak Amplitude (A) of each Audio block in the Expected Output files. 3. Calculate the RMSE of each Audio block by comparing the output (x) with the Expected Output files (y) Audio blocks. 4. Accept the AIM under test if, for each audio block, these the two conditions are satisfied: a. Data format of the Denoised Speech is the same with the Expected Output Files and b. RMSE < A* 0.1%

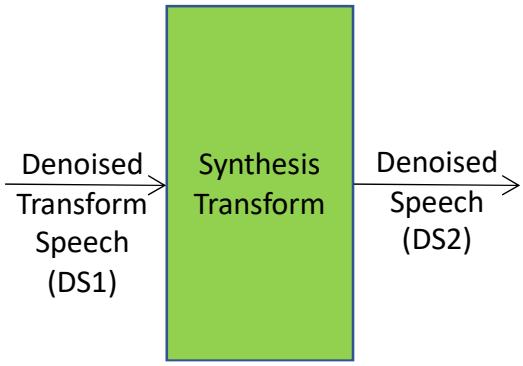

Figure 24 - Synthesis Transform Testing Flow

After the Tests, Conformance Tester shall fill out *Table 62Table 62*

Table 62 – Conformance Testing form of Enhanced Audioconference Experience (CAE-EAE) Synthesis Transform

<table>
<tr><td>Conformance Tester ID</td><td>Unique Conformance Tester Identifier assigned by MPAI</td></tr>
<tr><td>Standard, Use Case ID and Version</td><td>Standard ID and Use Case ID, Version and Profile of the standard in the form “CAE:EAE:1:0”.</td></tr>
<tr><td>Name of AIM</td><td>Synthesis Transform</td></tr>
<tr><td>Implementer ID</td><td>Unique Implementer Identifier assigned by MPAI Store.</td></tr>
<tr><td>AIM Implementation Version</td><td>Unique Implementation Identifier assigned by Implementer.</td></tr>
<tr><td>Neural Network Version*</td><td>Unique Neural Network Identifier assigned by Implementer.</td></tr>
<tr><td>Identifier of Conformance Testing Dataset</td><td>Unique Dataset Identifier assigned by MPAI Store.</td></tr>
<tr><td>Test ID</td><td>Unique Test Identifier assigned by Conformance Tester.</td></tr>
<tr><td>Actual output</td><td>The Conformance Tester will provide the following matrix with a limited number of input records (n) with the corresponding outputs. If an input record fails, the tester would specify the reason why the test case fails.
<table>
<tr><th>Input data (DS1)</th><th>Expected Output Data (DS2)</th><th>Data Format</th><th>RMSE</th></tr>
<tr><td>Denoised Transform Speech ID_1</td><td>Denoised Speech ID_1</td><td>T/F</td><td>< A* 0.1%</td></tr>
<tr><td>Denoised Transform Speech ID_2</td><td>Denoised Speech ID_2</td><td>T/F</td><td>< A* 0.1%</td></tr>
<tr><td>Denoised Transform Speech ID_3</td><td>Denoised Speech ID_3</td><td>T/F</td><td>< A* 0.1%</td></tr>
<tr><td>…</td><td>…</td><td></td><td>…</td></tr>
<tr><td>Denoised Transfom Speech ID_n</td><td>Denoised Speech ID_n</td><td>T/F</td><td>< A* 0.1%</td></tr>
</table></td></tr>
</table>

	Final evaluation: T/F Denoting with i, $0 \leq i < n = 50$, the record number in DS1 and DS2, the matrices reflect the results obtained with input records i with the corresponding outputs i. DS1 | DS2 | Synthesis Transform output value (obtained through the AIM under test) DS1[i] | DS2[i] | SynthesisTransform[i]
Execution time*	Duration of test execution.
Test comment*	Comments on test results and possible needed actions.
Test Date	yyyy/mm/dd.

** Optional field*

5.5.7 Enhanced Audioconference Experience (CAE-EAE) Packager

Table 63 gives the input/output data of the Packager.

Table 63 – I/O Data of Enhanced Audioconference Experience (CAE-EAE) Packager

AIM	Input Data	Output Data
Packager	Denoised Speech Microphone Array Geometry Audio Scene Geometry	Multichannel Audio + Audio Scene Geometry

Table 64 Table 64 gives the *Enhanced Audioconference Experience (CAE-EAE) Packager* Means and how they are used.

Table 64 – AIM Means and use of Enhanced Audioconference Experience (CAE-EAE) Packager

Means	Actions
Conformance Testing Dataset	DS1: *n* Test files including data in Denoised Speech format. DS2: *n* Microphone Array Geometry. DS3: *n* Audio Scene Geometry associated with the Denoised Speech. DS4: *n* Expected Output files including data in Multichannel Audio Stream format.
Procedure	1. Feed the AIM under test with the Test files (DS1, DS2, DS3). 2. Analyse the Multichannel Audio + Audio Scene Geometry with the Expected Output files (DS4).
Evaluation	1. Check the Multichannel Audio + Audio Scene Geometry data format with the given Expected Output files format. 2. Calculate the peak-to-peak Amplitude (A) of each Audio block in the Expected Output files. 3. Calculate the RMSE of each Audio block by comparing the output (x) with the Expected Output files (y) Audio blocks.

	4. Accept the AIM under test if, for each audio block, these the two conditions are satisfied: a. Data format of the Multichannel Audio + Audio Scene Geometry is the same as the Expected Output Files and b. RMSE < A * 0.1%

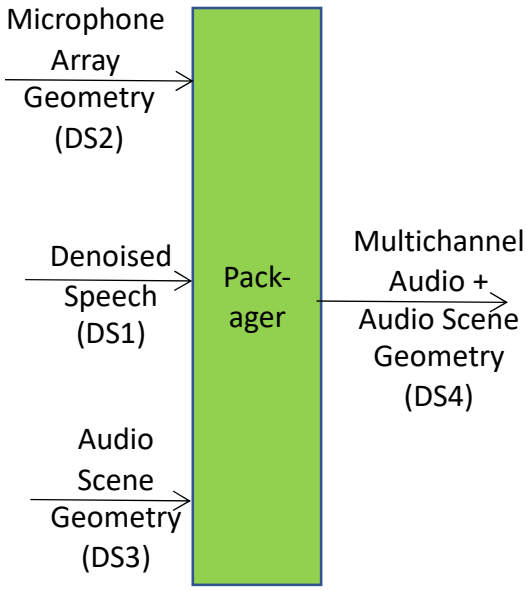

Figure 25 - Packager Testing Flow

After the Tests, Conformance Tester shall fill out *Table 65*.

Table 65 – Conformance Testing form of Enhanced Audioconference Experience (CAE-EAE) Packager

Conformance Tester ID	Unique Conformance Tester Identifier assigned by MPAI
Standard, Use Case ID and Version	Standard ID and Use Case ID, Version and Profile of the standard in the form "CAE:EAE:1:0".
Name of AIM	Packager
Implementer ID	Unique Implementer Identifier assigned by MPAI Store.
AIM Implementation Version	Unique Implementation Identifier assigned by Implementer.
Neural Network Version*	Unique Neural Network Identifier assigned by Implementer.
Identifier of Conformance Testing Dataset	Unique Dataset Identifier assigned by MPAI Store.
Test ID	Unique Test Identifier assigned by Conformance Tester.
Actual output	The Conformance Tester will provide the following matrix with a limited number of input records (*n*) with the corresponding outputs. If an input record fails, the tester would specify the reason why the test case fails.

Input data (DS1, DS2, DS3)	Expected Output Data (DS4)	Data Format	RMSE
Denoised Speech	Interleaved	T/F	< A *

ID_1 Microphone Array Geometry ID_1 Audio Scene Geometry ID_1	Multichannel Audio + Audio Scene Geometry ID_1		0.1%
Denoised Speech ID_2 Microphone Array Geometry ID_2 Audio Scene Geometry ID_2	Interleaved Multichannel Audio + Audio Scene Geometry ID_2	T/F	< A * 0.1%
Denoised Speech ID_3 Microphone Array Geometry ID_3 Audio Scene Geometry ID_3	Interleaved Multichannel Audio + Audio Scene Geometry ID_3	T/F	< A * 0.1%
…	…		…
Denoised Speech ID_n Microphone Array Geometry ID_n Audio Scene Geometry ID_n	Interleaved Multichannel Audio + Audio Scene Geometry ID_n	T/F	< A * 0.1%

Final evaluation : T/F

Denoting with i, $1 \leq i < n = 50$, the record number in DS1, DS2, and DS3, the matrices reflect the results obtained with input records i with the corresponding outputs i.

DS1	DS2	DS3	DS4	Packager output value (obtained through the AIM under test)
DS1[i]	DS2[i]	DS3[i]	DS4[i]	Packager[i]

Execution time*	Duration of test execution.
Test comment*	Comments on test results and possible needed actions.
Test Date	yyyy/mm/dd.

** Optional field*

Annex 1 – MPAI-wide terms and definitions (Normative)

The Terms used in this standard whose first letter is capital and are not already included in *Table 1* are defined in *Table 66Table 66.*

Table 66 – MPAI-wide Terms

Term	Definition
Access	Static or slowly changing data that are required by an application such as domain knowledge data, data models, etc.
AI Framework (AIF)	The environment where AIWs are executed.
AI Module (AIM)	A processing element receiving AIM-specific Inputs and producing AIM-specific Outputs according to according to its Function. An AIM may be an aggregation of AIMs.
AI Workflow (AIW)	A structured aggregation of AIMs implementing a Use Case receiving AIM-specific inputs and producing AIM-specific inputs according to its Function.
AIF Metadata	The data set describing the capabilities of an AIF set by the AIF Implementer.
AIM Metadata	The data set describing the capabilities of an AIM set by the AIM Implementer.
Application Programming Interface (API)	A software interface that allows two applications to talk to each other
Application Standard	An MPAI Standard specifying AIWs, AIMs, Topologies and Formats suitable for a particular application domain.
Channel	A physical or logical connection between an output Port of an AIM and an input Port of an AIM. The term “connection” is also used as a synonym.
Communication	The infrastructure that implements message passing between AIMs.
Component	One of the 9 AIF elements: Access, AI Module, AI Workflow, Communication, Controller, Internal Storage, Global Storage, MPAI Store, and User Agent.
Conformance	The attribute of an Implementation of being a correct technical Implementation of a Technical Specification.
Conformance Tester	An entity authorised by MPAI to Test the Conformance of an Implementation.
Conformance Testing	The process that verifies that the AIMs and/or the AIW of a Use Case part of a Standard have been implemented to produce data whose Semantics and Format conform with the Normative clauses of the relevant Use Case of the Technical Specification while providing a user experience level equal to or greater than the level specified in the Conformance Testing Specification.
Conformance Testing Means	Procedures, tools, data sets and/or data set characteristics to Test the Conformance of an Implementation.
Connection	A channel connecting an output port of an AIM and an input port of an AIM.
Controller	A Component that manages and controls the AIMs in the AIF, so that they execute in the correct order and at the time when they are needed.
Data	Information in digital form.
Data Format	The standard digital representation of Data.
Data Semantics	The meaning of Data.
Device	A hardware and/or software entity running at least one instance of an AIF.

Ecosystem	The ensemble of the following actors: MPAI, MPAI Store, Implementers, Conformance Testers, Performance Testers and Users of MPAI-AIF Implementations as needed to enable an Interoperability Level.
Event	An occurrence acted on by an Implementation.
Explainability	The ability to trace the output of an Implementation back to the inputs that have produced it.
Fairness	The attribute of an Implementation whose extent of applicability can be assessed by making the training set and/or network open to testing for bias and unanticipated results.
Function	The operations effected by an AIW or an AIM on input data.
Global Storage	A Component to store data shared by AIMs.
Identifier	A name that uniquely identifies an Implementation.
Implementation	1. An embodiment of the MPAI-AIF Technical Specification, or 2. An AIW or AIM of a particular Level (1-2-3).
Internal Storage	A Component to store data of the individual AIMs.
Interoperability	The ability to functionally replace an AIM/AIW with another AIM/AIW having the same Interoperability Level
Interoperability Level	The attribute of an AIW and its AIMs to be executable in an AIF Implementation and to be: 1. Implementer-specific and satisfying the MPAI-AIF Standard *(Level 1)*. 2. Specified by an MPAI Application Standard (*Level 2)*. 3. Specified by an MPAI Application Standard and certified by a Performance Assessor (*Level 3)*.
Knowledge Base	Structured and/or unstructured information made accessible to AIMs via MPAI-specified interfaces
Message	A sequence of Records.
Normativity	The set of attributes of a technology or a set of technologies specified by the applicable parts of an MPAI standard.
Performance	The attribute of an Implementation of being Reliable, Robust, Fair and Replicable.
Performance Assessment	The normative document specifying the procedures, the tools, the data sets and/or the data set characteristics to Assess the Grade of Performance of an Implementation.
Performance Assessment Means	Procedures, tools, data sets and/or data set characteristics to Assess the Performance of an Implementation.
Performance Assessor	An entity authorised by MPAI to Assess the Performance of an Implementation in a given Application domain
Port	A physical or logical communication interface of an AIM.
Profile	A particular subset of the technologies used in MPAI-AIF or an AIW of an Application Standard and, where applicable, the classes, other subsets, options and parameters relevant to that subset.
Record	Data with a specified structure.
Reference Model	The AIMs and theirs Connections in an AIW.
Reference Software	A technically correct software implementation of a Technical Specification containing source code, or source and compiled code.
Reliability	The attribute of an Implementation that performs as specified by the Application Standard, profile and version the Implementation refers to, e.g., within the application scope, stated limitations, and for the period of time specified by the Implementer.

Replicability	The attribute of an Implementation whose Performance, as Assessed by a Performance Assessor, can be replicated, within an agreed level, by another Performance Assessor.
Robustness	The attribute of an Implementation that copes with data outside of the stated application scope with an estimated degree of confidence.
Scope	The domain of applicability of an MPAI Application Standard
Service Provider	An entrepreneur who offers an Implementation as a service (e.g., a recommendation service) to Users.
Specification	A collection of normative clauses.
Standard	The ensemble of Technical Specification, Reference Software, Conformance Testing and Performance Assessment of an MPAI application Standard.
Technical Specification	(Framework) the normative specification of the AIF. (Application) the normative specification of the set of AIWs belonging to an application domain along with the AIMs required to Implement the AIWs that includes: 1. The formats of the Input/Output data of the AIWs implementing the AIWs. 2. The Connections of the AIMs of the AIW. 3. The formats of the Input/Output data of the AIMs part of the AIW.
Testing Laboratory	A laboratory accredited by MPAI to Assess the Grade of Performance of Implementations.
Time Base	The protocol specifying how Components can access timing information
Topology	The set of AIM Connections of an AIW.
Use Case	A particular instance of the Application domain target of an Application Standard.
User	A user of an Implementation.
User Agent	The Component interfacing the user with an AIF through the Controller
Version	A revision or extension of a Standard or of one of its elements.
Wrapper AIM	An AIM implementation performing the function of an MPAI standardised AIM by calling an openly accessible, possibly paid internet service provided by a third party. E.g., Microsoft TrueText.
Zero Trust	A cybersecurity model primarily focused on data and service protection that assumes no implicit trust.

Annex 2 - Notices and Disclaimers Concerning MPAI Standards (Informative)

The notices and legal disclaimers given below shall be borne in mind when downloading and using approved MPAI Standards.

In the following, "Standard" means the collection of four MPAI-approved and published documents: "Technical Specification", "Reference Software" and "Conformance Testing" and, where applicable, "Performance Testing".

Life cycle of MPAI Standards

MPAI Standards are developed in accordance with the MPAI Statutes. An MPAI Standard may only be developed when a Framework Licence has been adopted. MPAI Standards are developed by especially established MPAI Development Committees who operate on the basis of consensus, as specified in Annex 1 of the MPAI Statutes. While the MPAI General Assembly and the Board of Directors administer the process of the said Annex 1, MPAI does not independently evaluate, test, or verify the accuracy of any of the information or the suitability of any of the technology choices made in its Standards.

MPAI Standards may be modified at any time by corrigenda or new editions. A new edition, however, may not necessarily replace an existing MPAI standard. Visit the web page to determine the status of any given published MPAI Standard.

Comments on MPAI Standards are welcome from any interested parties, whether MPAI members or not. Comments shall mandatorily include the name and the version of the MPAI Standard and, if applicable, the specific page or line the comment applies to. Comments should be sent to the MPAI Secretariat. Comments will be reviewed by the appropriate committee for their technical relevance. However, MPAI does not provide interpretation, consulting information, or advice on MPAI Standards. Interested parties are invited to join MPAI so that they can attend the relevant Development Committees.

Coverage and Applicability of MPAI Standards

MPAI makes no warranties or representations of any kind concerning its Standards, and expressly disclaims all warranties, expressed or implied, concerning any of its Standards, including but not limited to the warranties of merchantability, fitness for a particular purpose, non-infringement etc. MPAI Standards are supplied "AS IS".

The existence of an MPAI Standard does not imply that there are no other ways to produce and distribute products and services in the scope of the Standard. Technical progress may render the technologies included in the MPAI Standard obsolete by the time the Standard is used, especially in a field as dynamic as AI. Therefore, those looking for standards in the Data Compression by Artificial Intelligence area should carefully assess the suitability of MPAI Standards for their needs.

IN NO EVENT SHALL MPAI 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

Annex 3 – The Governance of the MPAI Ecosystem (Informative)

Level 1 Interoperability

With reference to **Error! Reference source not found.**, MPAI issues and maintains a standard – called MPAI-AIF – whose components are:

1. An environment called AI Framework (AIF) running AI Workflows (AIW) composed of interconnected AI Modules (AIM) exposing standard interfaces.
2. A distribution system of AIW and AIM Implementation called MPAI Store from which an AIF Implementation can download AIWs and AIMs.

Implementers' benefits	Upload to the MPAI Store and have globally distributed Implementations of - AIFs conforming to MPAI-AIF. - AIWs and AIMs performing proprietary functions executable in AIF.
Users' benefits	Rely on Implementations that have been tested for security.
MPAI Store's role	- Tests the Conformance of Implementations to MPAI-AIF. - Verifies Implementations' security, e.g., absence of malware. - Indicates unambiguously that Implementations are Level 1.

Level 2 Interoperability

In a Level 2 Implementation, the AIW shall be an Implementation of an MPAI Use Case and the AIMs shall conform with an MPAI Application Standard.

Implementers' benefits	Upload to the MPAI Store and have globally distributed Implementations of - AIFs conforming to MPAI-AIF. - AIWs and AIMs conforming to MPAI Application Standards.
Users' benefits	- Rely on Implementations of AIWs and AIMs whose Functions have been reviewed during standardisation. - Have a degree of Explainability of the AIW operation because the AIM Functions and the data Formats are known.
Market's benefits	- Open AIW and AIM markets foster competition leading to better products. - Competition of AIW and AIM Implementations fosters AI innovation.
MPAI Store's role	- Tests Conformance of Implementations with the relevant MPAI Standard. - Verifies Implementations' security. - Indicates unambiguously that Implementations are Level 2.

Level 3 Interoperability

MPAI does not generally set standards on how and with what data an AIM should be trained. This is an important differentiator that promotes competition leading to better solutions. However, the performance of an AIM is typically higher if the data used for training are in greater quantity and more in tune with the scope. Training data that have large variety and cover the spectrum of all cases of interest in breadth and depth typically lead to Implementations of higher "quality".

For Level 3, MPAI normatively specifies the process, the tools and the data or the characteristics of the data to be used to Assess the Grade of Performance of an AIM or an AIW.

Implementers' benefits	May claim their Implementations have passed Performance Assessment.
Users' benefits	Get assurance that the Implementation being used performs correctly, e.g., it has been properly trained.
Market's benefits	Implementations' Performance Grades stimulate the development of more Performing AIM and AIW Implementations.

MPAI Store's role	-	Verifies the Implementations' security
	-	Indicates unambiguously that Implementations are Level 3.

The MPAI ecosystem

The following *Figure 26* is a high-level description of the MPAI ecosystem operation applicable to fully conforming MPAI implementations:

1. MPAI establishes and controls the not-for-profit MPAI Store.
2. MPAI appoints Performance Assessors.
3. MPAI publishes Standards.
4. Implementers submit Implementations to Performance Assessors.
5. If the Implementation Performance is acceptable, Performance Assessors inform Implementers (step 5a) and MPAI Store.
6. Implementers submit Implementations to the MPAI Store tested for Conformance and security.
7. Users download and use Implementations and submit experience scores.

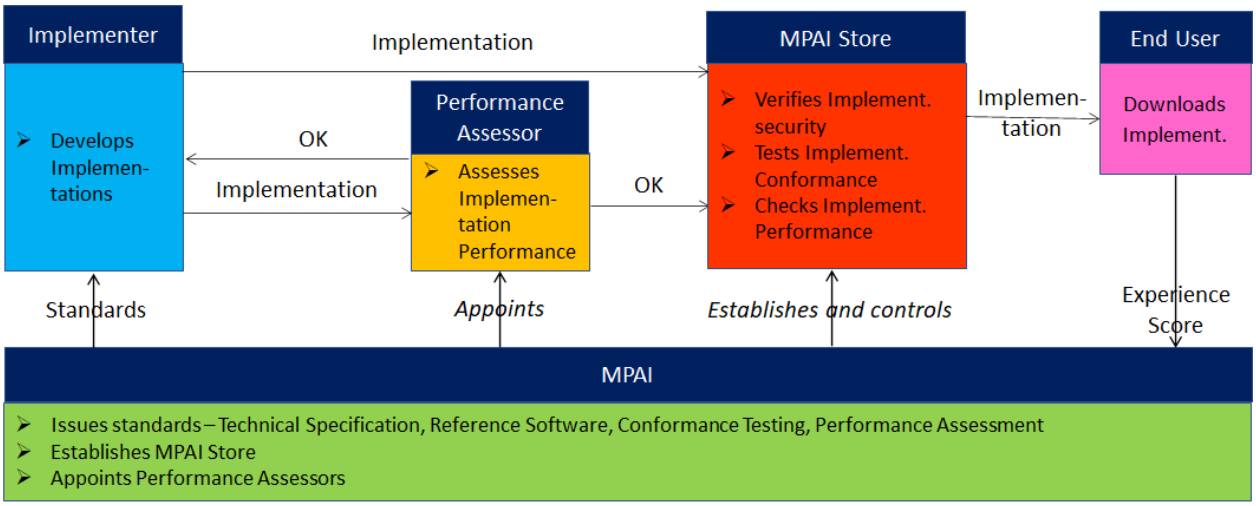

Figure 26 - The MPAI ecosystem operation.