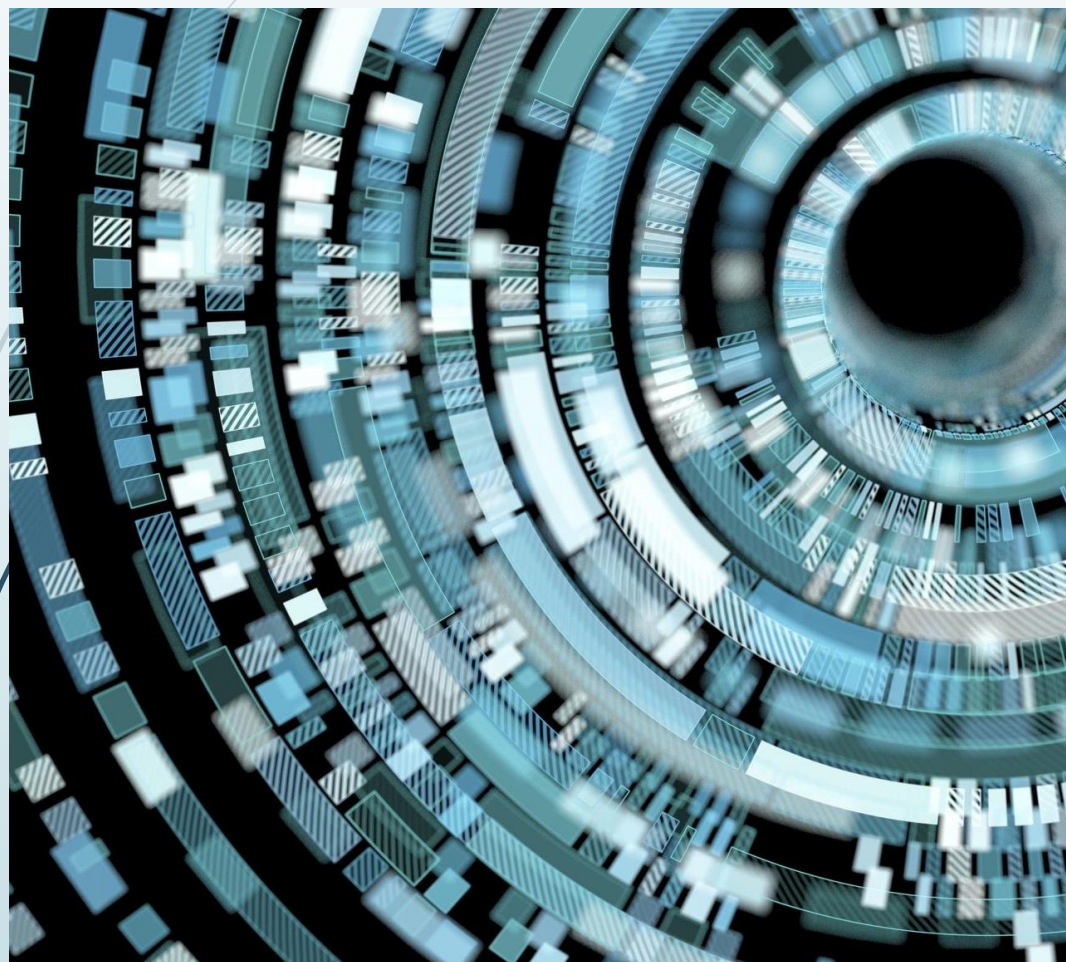




# MPAI Call for technologies Six Degrees of Freedom Audio (CAE-6DF)

Leonardo Chiariglione – MPAI

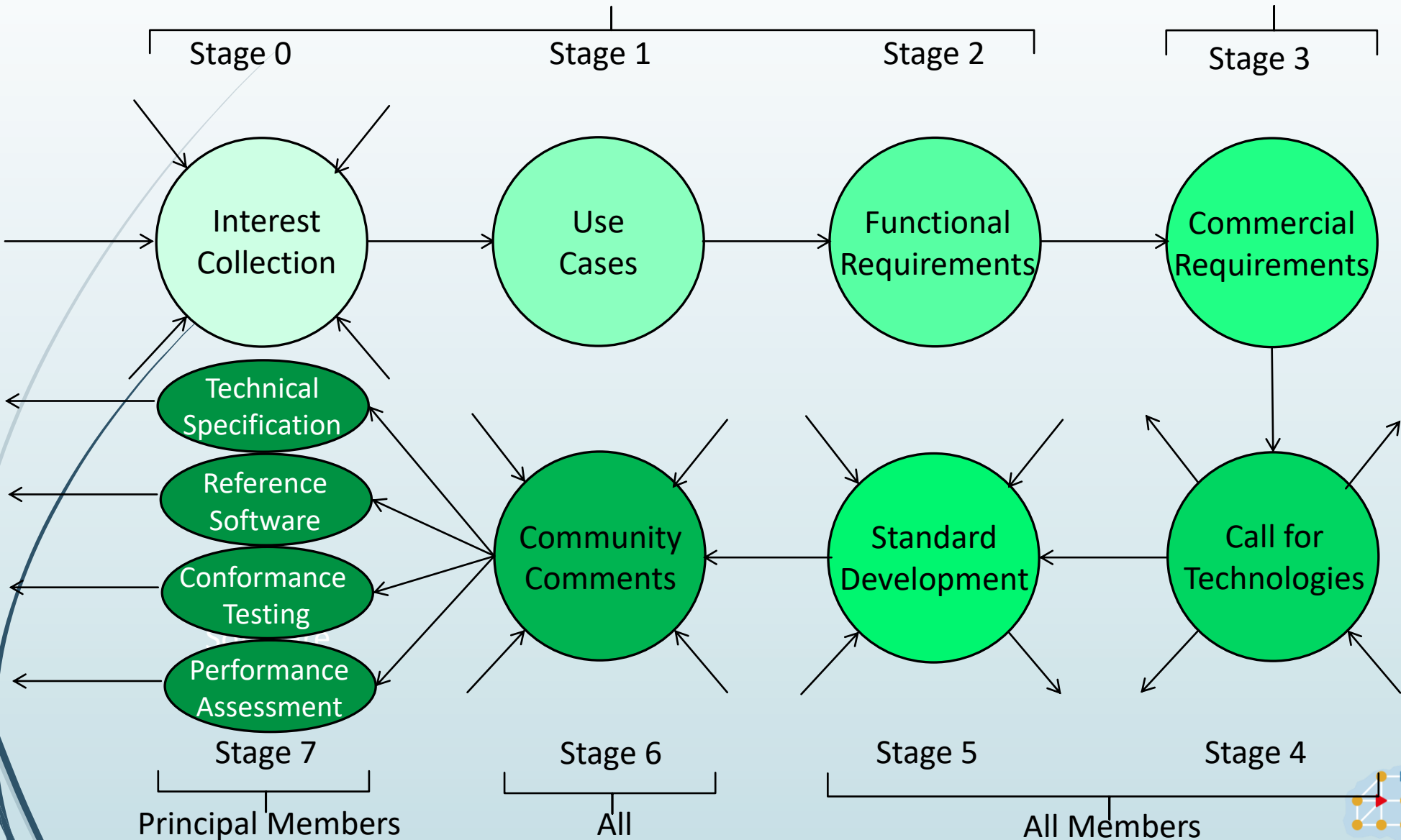
# Two words about MPAI



- MPAI – Moving Picture, Audio, and Data Coding by Artificial Intelligence.
- What is is: international, unaffiliated, non-profit standard developing organization.
- What it does: standards for AI-enabled data coding with clear IP guidelines.
- Portfolio of 11 standards and about to publish a new one.
- Five standards adopted by IEEE as IEEE standards and more to come



# Rigorous standard development process



# An overview of approved MPAI standards

Category	Title	Acronym
System	AI Framework	MPAI-AIF
Audio	Context-based Audio Enhancement	MPAI-CAE
Automotive	Connected Autonomous Vehicle	MPAI-CAV
Finance	Compression and Understanding of Industrial Data	MPAI-CUI
System	Governance of the MPAI Ecosystem	MPAI-GME
Human-machine	Human and Machine Communication	MPAI-HMC
Human-machine	Multimodal Conversation	MPAI-MMC
Metaverse	MPAI Metaverse Model	MPAI-MMM
Watermarking	Neural Network Watermarking	MPAI-NNW
Objects & Scenes	Object and Scene Description	MPAI-OSD
Avatars	Portable Avatar Format	MPAI-PAF

# On the way to become standards

Category	Title	Acronym
Health	AI for Health	MPAI-AIH
Video	End-to-End Video Coding	MPAI-EEV
Video	Up-sampling Filter for Video Application	MPAI-UFV
Online Gaming	Server-based Predictive Multiplayer Gaming	MPAI-SPG
System	AI Module Profiles	MPAI-PRF
Entertainment	XR Venues – Live Theatrical Performance	MPAI-XRV

MPAI is launching a new project: **Six Degrees of Freedom (6DF) Audio**, including:

Call for Technologies

Use Cases and Functional Requirements

Framework Licence

Template for Responses

# Call for Technologies/1

- This Call for Technologies: Context-based Audio Enhancement (MPAI-CAE) – Six Degrees of Freedom Audio (CAE-6DF) **invites any party** able and wishing to contribute to the development of the planned MPAI CAE-6DF Technical Specification **to submit a response**.
- If they own technologies relevant to this Call, they are required to eventually license their technologies according to CAE-6DF Framework Licence if their technologies are selected by MPAI for possible modification and inclusion in the planned CAE-6DF Technical Specification.
- Any respondent who is not an MPAI member and wishes to participate in the development of CAE-6DF shall join MPAI. If they own accepted technologies and do not join MPAI, they lose the opportunity to have their technologies included in the planned CAE-6DF.

# Call for Technologies/1

- The scope of this Call is restricted to responses whose specification in the planned CAE-6DF conforms to Use Cases and Functional Requirements.
- Respondents are welcome to additionally make:
  - Comments on any technical element of [4].
  - Technologies based on motivated proposals to extend or add new functional requirements provided they:
    - Are in line with the scope of the functional requirements.
    - Satisfy the Framework Licence.

# Dates

Step	Date	Time
CAE-6DF Call for Technologies issued.	2024/05/15	17:00 UTC
CAE-6DF Call for Technologies presented online.	2024/05/28	16:00 UTC
Notification of intention to submit proposal (Annex A).	2024/06/04	23:59 UTC
Response submission deadline.	2024/09/16	23:59 UTC
Start of response evaluation.	2024/09/30	(MPAI-48)





MPAI Context-based Audio Enhancement (MPAI-CAE)  
**Call for Technologies:  
Six Degrees of Freedom (CAE-6DF)**



Marina Bosi, Chair MPAI-CAE  
CCRMA, Stanford University

# Scope of CAE-6DF

## ➤ Goal:

- To develop Technical Specifications (CAE-6DF TS) for immersive audio that shall enable **realistic auditory experiences in virtual environments, responding to locomotive/translational motion of users**

## ➤ Key Aspects:

- Representation of both static and dynamic (i.e., moving) audio entities
- Provisions for integration with visual elements (e.g., volumetric visual content) to support audiovisual immersion
- Low latency and high responsiveness

# Use Cases Overview

## ➤ Immersive Content

- Enhances the sense of **presence** and **realism**
- Allow users to interact naturally within the virtual environment
- Diverse applications in, for example, **entertainment** and **education**:
  - Examples: concerts, radio dramas, lectures, opera/theatre/dance performances, documentaries

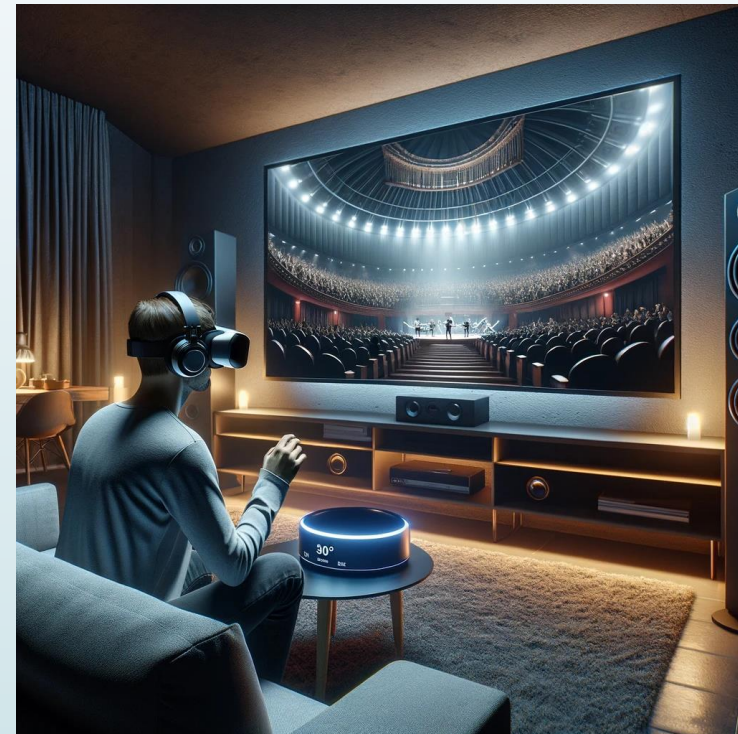
# Use Case 1

## Immersive Concert Experience

User experiences a concert as if they are in a concert hall

### Key Features:

- Immersive audio rendering replicates the concert hall acoustics
- Spatial and temporal synchronization with 360° video
- User movement within the living room corresponds to changes in the audio experience



*This image was generated with the assistance of AI.*

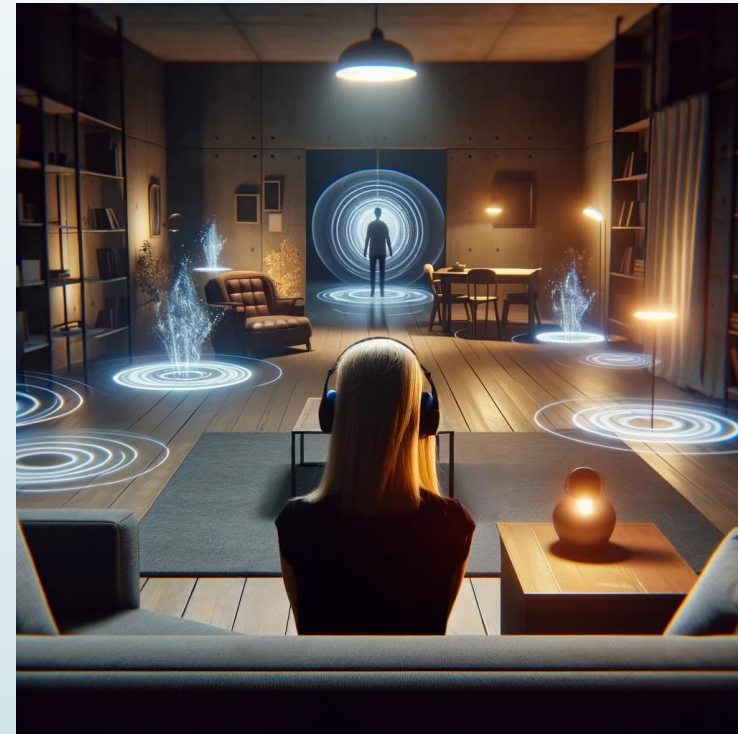
# Use Case 2

## Immersive Radio Drama

User experiences a radio drama with immersive audio

### Key Features:

- Ability to walk or virtually teleport within the rendering space
- User-selected auditory perspectives enhance the storytelling experience



*This image was generated with the assistance of AI.*

# Use Case 3

## Virtual Lecture

User uses an immersive audio-visual display to attend a virtual lecture

### Key Features:

- Audio to accompany 360° videos captured at different points in a lecture theatre
- User can physically walk or be virtually teleported to different perspectives
- Synchronization with 360° video enhances the learning experience



*This image was generated with the assistance of AI.*

# Use Case 4

## Immersive Theatre Experience

User experiences opera, ballet, dance, or theatre in 360° or 6DoF

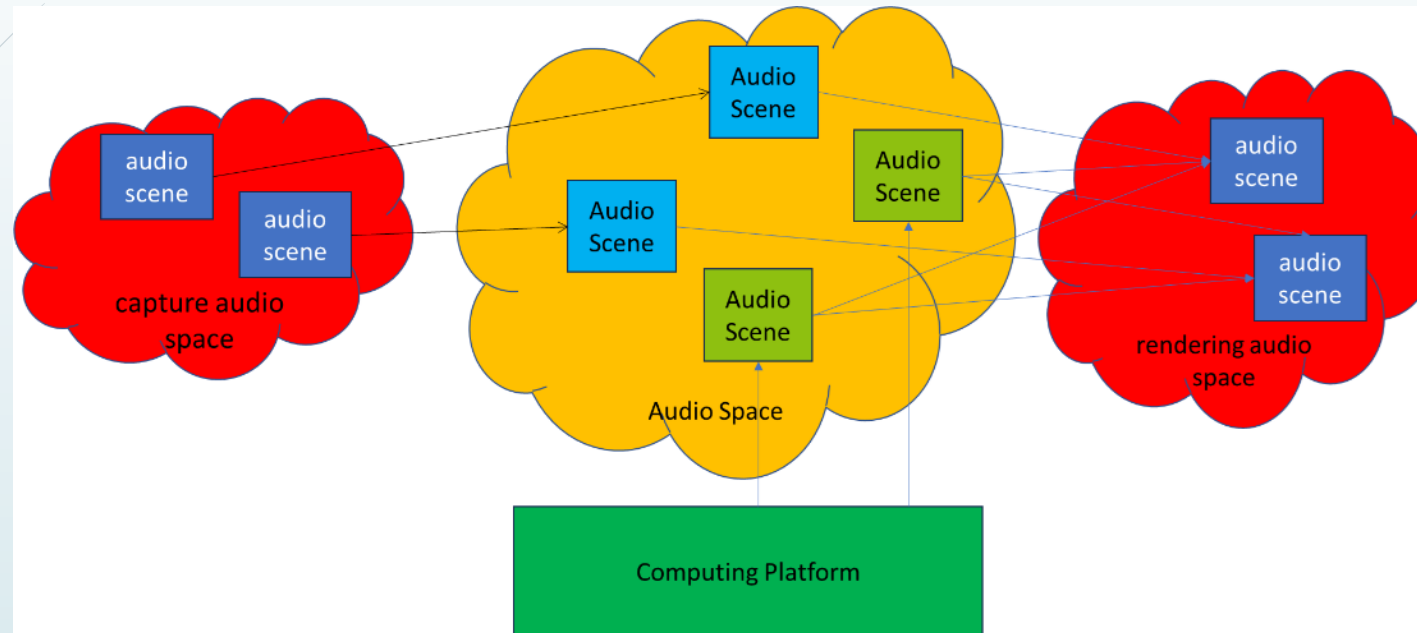
### Key Features:

- Move within the living room to experience different perspectives (e.g., on stage, different seats)
- Immersive audio recorded during an actual performance brings user-defined listening perspectives



*This image was generated with the assistance of AI.*

# CAE-6DF Concept



- The *audio scenes* (e.g., a concert) captured in a *capture audio space* (e.g., a concert hall, recording studio) are converted to **CAE-6DF Audio Scenes** in **CAE-6DF Audio Space**
- **CAE-6DF Audio Scenes** in **CAE-6DF Audio Space** are rendered as *audio scenes* in a *rendering audio space* (e.g., a user's living room)



# Functional Requirements Overview

- Ensure high-quality audio experience in immersive environments
- Allow user's translational and rotational movement within the rendered audio scene at interactive update rates
- Allow for audio-visual synchronization for coherent experiences
- Model and respond to effects of locomotion on human audio-visual perception
- Provide auditory distance/depth perception for realistic interactions

# Audio-Visual

Alignment of audio scenes with visual scenes

**The technology shall...**

- Ensure a coherent and immersive experience
- Enhance the realism of virtual environments

# Locomotion/Orientation Response

Impact of user's movement on audio-visual perception

## The technology shall...

- Provide realistic auditory and visual feedback
- Allow for a natural interaction within the virtual environment

# Distance Perception

Visual and auditory modalities affecting each other

## The technology shall...

- Facilitate accurate perception of distance in the virtual environment
- Enhance the realism of the audio experience via accurate depth perception

# Content Profiles

## 1. Scene-based:

- Accurate reconstruction of the audio scene
- High degree of correspondence to the original acoustic environment

## 2. Object-based:

- Audio Objects and metadata for perceptual representation
- Allows for creative and flexible audio experiences

## 3. Hybrid:

- Combination of scene-based and object-based content
- Enhances the richness and depth of the audio experience

# Rendering Modalities

## Loudspeaker-based:

- Content delivered through (*at least two but potentially more*) loudspeakers
- Provides a room-filling audio experience



## Headphone-based:

- Content delivered through headphones
- Ensures immersive audio experience without external disturbances



# Rendering Space Characteristics

## Shape and Dimensions:

Not larger than the captured space to ensure accurate reproduction

## Acoustic Characteristics:

- Early decay time, frequency mode density, echo density, reverberation time, energy decay curve, and background noise
- Ensure the rendered space closely matches the original environment

# User Movement

## Types:

- Actual locomotion/orientation tracked by sensors
- Virtual locomotion/orientation controlled by VR controllers

## The technology shall...

- Accomodate means for real-time tracking and respond to user movements at interactive rates
- Enhance the sense of presence and interaction within the virtual environment following user's movement



# Latency Requirements

## Ideal Range of Latency:

20 ms (up to 50 ms) from movement to rendering

## The technology shall...

- Ensure high responsiveness to user movements
- Minimize delay to enhance the immersive experience

# Summary

**Technologies submitted to the CfT for CAE-6DF shall:**

- ▶ Provide full auditory immersion
- ▶ Be compatible with immersive visual content and state-of-the-art VR systems and controllers
- ▶ Cater to at least one (and ideally all) of the four described use cases
- ▶ Provide loudspeaker and headphone-based rendering capabilities
- ▶ Provide low latency (20-50 ms) and be highly responsive

# CAE 6DF Call For Technologies Documents

Call for Technologies:  
6DoF Audio (CAE-6DF)  
(N1763)

Functional Requirements:  
6DoF Audio (CAE-6DF)  
(N1764)

Framework License: 6DoF  
Audio (CAE-6DF) (N1765)

Template for  
Responses: 6DoF Audio  
(CAE-6DF) (N1766)

# What is next?



<b>CAE-6DF Call for Technologies issued</b>	<b>2024/05/15</b>	<b>17:00 UTC</b>
<b>CAE-6DF Call for Technologies presented online</b>	2024/05/28	16:00 UTC
<b>Notification of intention to submit proposal</b>	2024/06/04	23:59 UTC
<b>Response submission deadline</b>	2024/09/16	23:59 UTC
<b>Start of response evaluation</b>	2024/09/30	(MPAI-48)



The Context-based  
Audio Enhancement  
Development

Committee  
(CAE-DC)

Marina Bosi, Chair  
MPAI-CAE

THANK YOU

# MPAI-6DF Framework Licence - Coverage

- ▶ This Framework Licence applies to Technical Specification: Context-based Audio Enhancement (MPAI-CAE) – Six-DoF Audio (CAE-6DF) (in the following “Standard”) as approved by Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI).
- ▶ The Standard may include technologies that have been standardised by other MPAI Technical Specifications. Use of the Essential IPR of the Standard will be based on the Conditions of use defined in Section 3.

# Conditions of use of the Essential IPR's/1

The Standard Essential IPR holders commit themselves to issue a Licence with the following conditions:

1. The Licence will be in compliance with generally accepted principles of competition law and the MPAI Statutes.
2. The Licence will cover:
  1. All claims to Essential IPR and copyright of a Licensor that are practised by a Licensee of the Standard.
  2. Development Rights and Implementation Rights.
3. The Licence will grant access to Essential IPRs of the Standard in a non-discriminatory fashion.

# Conditions of use of the Essential IPR's/2

4. The scope of the Licence may be subject to legal, bias, ethical and moral limitations.
5. Royalties will apply:
  5. To all profiles of the Standard including the Baseline Profile.
  6. To any Implementation that is based on the Standard, with the exclusion of the types of implementations specified in clause 7.1.
  7. On a worldwide basis.
6. An Implementation of the Standard may use other IPR to extend the Implementation Standard or to provide additional functionalities.



# Conditions of use of the Essential IPR's/3

## 7. Exemptions:

1. A Licence for Development and Implementation Rights, to the extent it is developed and implemented only for the purpose of evaluation or demo solutions or technical trials, will be free of charge.
  2. A Licence may be granted free of charge for particular uses if so decided by the licensors.
  3. A free of charge Licence for a limited time and a limited amount of forfeited royalties will be granted on request.
8. Holders of Standard Essential Patents (SEP) will express a preference on the entity that should administer the patent pool of their SEPs.

# Conditions of use of the Essential IPR's/4

9. The Licence will be issued before commercial implementations of the Standard become available on the market. Commercial implementation implies General Availability to any users and does not include trials.
10. The total cost of the Licences issued by IPR holders will be in line with the total cost of the Licences for similar technologies standardised in the context of Standard Development Organisations.
11. The total cost of the Licences will consider the market value of the SEPs.

# Template for Responses

- ▶ This document is provided as a help to those who intend to submit a response to Call for Technologies: Template for Responses - Context-based Audio Enhancement (MPAI-CAE) – Six-DoF Audio (CAE-6DF).
- ▶ It contains mandatory text that a proposal shall include.



We look forward to working  
**with you**  
on this exciting MPAI project!

<https://mpai.community/>

Join MPAI  
Share the fun  
Build the future



**MPAI.**  
community