

# 사실 표준화 MPAI 그룹의 표준화 동향

최미란 한국전자통신연구원 언어지능연구실 책임연구원/표준전문위원

## 1. 머리말

빠르게 성장하고 있는 인공지능 시대의 흐름에 맞춰서 새롭게 창시된 사실 표준화 그룹인 MPAI(Moving Picture, Audio and Data Coding by Artificial Intelligence) 그룹(<https://mpai.community>)은 MPEG의 공동 창시자인 이탈리아 공학자 레오나르도 키아릴리오네(Leonardo Chiariglione)에 의해 제안돼 수 개월간의 준비 기간을 거친 뒤, 2020년 9월 30일에 개최된 창립총회에서 공식적인 표준화 활동의 시작을 선언했다. MPAI는 제네바에 기반을 둔 국제 비영리 단체로 현재 15개국의 33개 기관이 회원으로 활동한다. 한국에서는 한국전자통신연구원, 경희대학교, 산업계의 클레온(KLleon) 회사가 회원으로 가입해 MPAI 운영과 표준화 개발에 활발히 참여하고 있다.

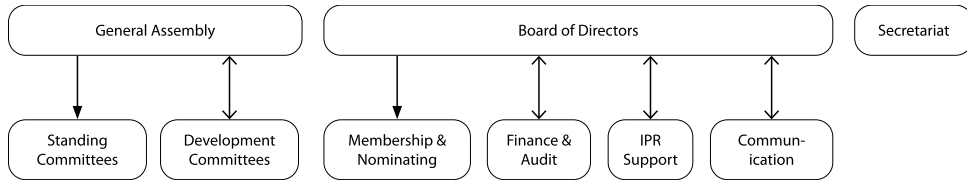
MPAI는 인공지능을 핵심기술로 하는 데이터 코딩 표준을 개발하고자 설립된 첫 번째 표준단체다. 여기에서 데이터 코딩이란 일정한 형식으

로 표현된 데이터를 특정한 응용에 좀 더 적당한 표현 방식으로 변환시키는 방법을 의미한다. 이렇게 변형된 형태의 데이터는 가능한 최대로 데이터의 본래의 의미를 유지해야 하며 특정 응용에서 중요한 의미의 양상을 보여줄 수 있어야 한다.

사실표준화 단체가 ISO나 ITU-T와 같은 공적표준화 단체와 다른 점은 시급한 기술을 바로 사용할 수 있도록 표준을 빠르게 개발해 개발자들이 시스템에 표준을 활용할 수 있도록 한다는 것이다. 그렇기에 MPAI 참여 그룹에는 미국 실리콘밸리의 벤처기업들을 포함하여 각국 산업계의 회원이 다수 포함돼 있다. 본고에서는 이렇게 빨리 진행되는 MPAI 표준화의 현황을 소개하기 위해 MPAI의 표준 진행 절차와 현재 진행되고 있는 표준의 개발 내용을 소개하고 향후 전개될 표준화를 전망해 본다.

## 2. MPAI의 표준개발 체계

공식적으로 2020년 9월 30일에 설립된



[그림 1] MPAI 조직 구성

MPAI는 [그림 1]과 같은 조직체계를 갖췄다. 최상위 운영그룹인 5인 구성 이사회에 한국도 이사로 참여해 MPAI 운영에 아시아 대표로서 의사결정에 중요한 역할을 하고 있다. 표준화 활동을 살펴보면 MPAI 설립 2개월 전부터 각 분야의 인공지능 전문가들이 온라인 회의를 통해 MPAI 표준이 기여할 수 있는 분야의 유즈케이스의 개발을 진행했다. 일반적으로 MPAI는 다른 사실 표준화 그룹과 마찬가지로 국제 공동 협력에 의해 MPAI의 임무를 준수하고 실행 목표를 가지고 관련 당사자들이 회의를 통해 표준을 개발하는 개방 단체라고 생각할 수 있다.

MPAI의 표준개발 체계는 전문가들과 설립 후 바로 논의됐으며 다음과 같은 내용의 표준개발 단계의 구성이 결정됐다.

- **1단계 유즈케이스(UC, Use Case):** 이 단계에서는 회원을 포함해 누구나 유즈케이스를 제안할 수 있으며 산업계에 적용하기에 적합한 핵심 분야로 지정돼 수집되고 분류된다.

- **2단계 기능요구사항(FR, Functional requirements):** 표준이 지원해야 할 기능적인 요구사항을 파악하고 문서화하는 단계다. 이 단계와 이전 단계에서는 관련 회원의 동의하에 비회원도 MPAI 회의에 참여할 수 있다.
- **3단계 상업요구사항(CR, Commercial Requirements):** 관련 표준에 대한 프레임워크 라이선스가 개발되는 단계로 표준에 기술을 제안하려는 사람들은 이 라이선스를 승인해야 한다.
- **4단계 기술공모(CT, Call for Technologies):** 기능적인 요구사항과 상업요구사항을 모두 만족시키는 기술의 공모가 발표되는 단계로 제안자들은 표준이 승인된 후에 프레임워크 라이선스에 따를 것을 선언해야 한다.
- **5단계 표준개발(SD, Standard Development):** 이 단계에서는 표준 개발을 위한 위원회가 구성되어 합의에 의한 표준 개발이 이뤄진다.
- **6단계 MPAI 표준(MS, MPAI Standard):** 이 단계는 표준이 완성되는 단계로 MPAI 회원들은 라이선스 동의서를 선언하고 비회원은 MPAI 표준을 사용하기 위한 동의서에 서명한 후에 사용할 수 있다.

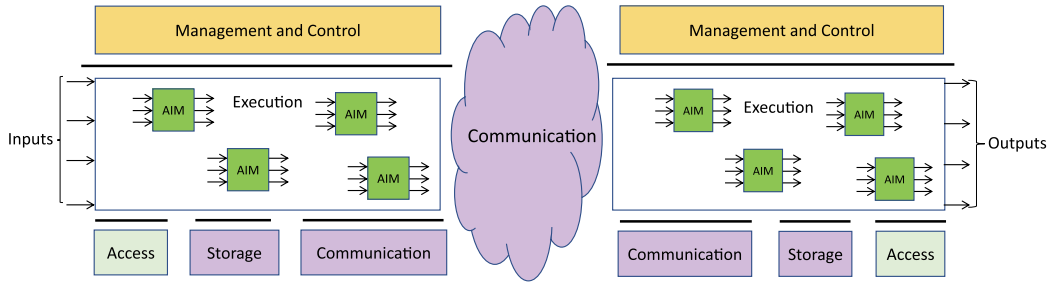
위와 같은 단계가 규정됐으며 현재 오디오, 비디오, 자연어처리 관련 응용 분야 등 여러 개의 표준 후보가 해당 단계에서 개발되고 있다.

## 2. MPAI의 표준 개발 현황

본 장에서는 현재 MPAI에서 개발되고 있는

<표 1> MPAI 표준 목록

표준 약자	표준 제목	표준 단계
MPAI-AIF	AI Framework (AI 프레임워크)	SD
MPAI-CAE	Context-based Audio Enhancement (맥락 기반의 오디오 품질 개선)	SD
MPAI-MMC	Multi-Modal Conversation (멀티모달 대화)	SD
MPAI-CUI	Compression and Understanding of Industrial Data (산업 데이터의 압축과 이해)	SD
MPAI-GSA	Integrative Analysis of Genomic/Sensor Experiments (게놈과 센서 실험을 위한 통합 분석)	FR
MPAI-SPG	Server-based Predictive Multiplayer Gaming (서버 기반의 예측 가능 다인 참여 게임)	FR
MPAI-EVC	AI-Enhanced Video Coding (인공지능에 의한 향상된 비디오 코딩)	UC



[그림 2] MPAI-AI 프레임워크 (AIF)[1]

표준의 소개와 개발 현황에 대해 논의한다. 개발 표준의 목록과 각 해당 표준 단계는 <표 1>과 같다.

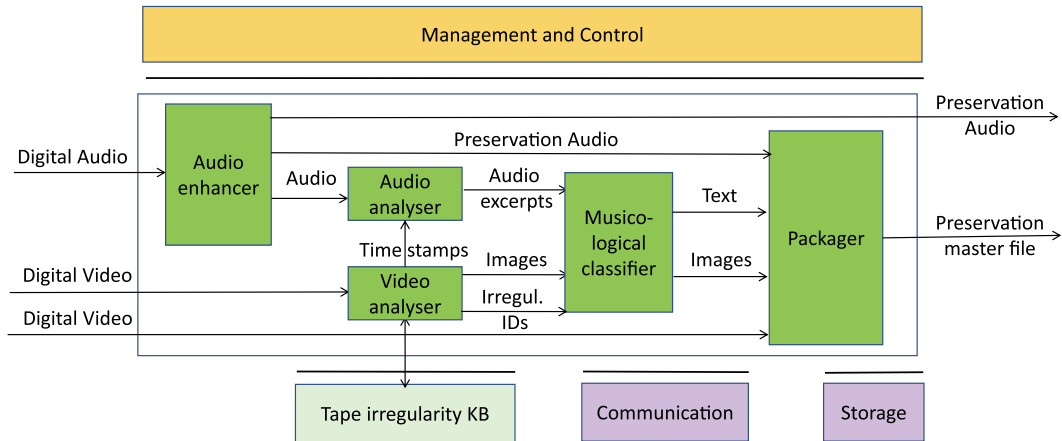
## 2.1 MPAI-AIF: MPAI 유즈케이스를 지원하기 위한 표준 프레임워크

MPAI AI Framework (AIF)는 MPAI에서 개발되는 모든 표준의 기반이 되는 프레임워크 표준으로 2020년 12월 16일에 기술공모가 발표됐고[1] 다음 해 2월까지 기술 제안이 이뤄졌다. 현재 MPAI는 제안된 기술들을 검토해 2021년 7월에 표준 완성을 목표로 프레임워크 표준을 개발하고 있다. 현재 구성된 MPAI-AIF 참조 모델은 [그림 2]와 같이 정의된다. MPAI-AIF 표준의 범위는 하드웨어와 소프트웨어에 장착된 인공지능과 데이터 처리 기술의 용이한 상호 연결과 상호 운용을 가능하게 하는 프레임워크를 제공하는 것이다. AI 프레임워크 환경에서 이러한 기술들은 AI 모듈(AIMs)이라고 부르는 표준 인터페이스를 사용하는 모듈에 캡슐화되는 것이다. MPAI-AIF는 구성 요소 기반 개발 방식의 원칙을 수용하여 독립적인 모듈들인 AIM을 시스템에 재사용할 수 있도록 하는 방식을 채택했다.

## 2.2 MPAI-CAE (Context-based Audio

### Enhancement): 맥락 기반의 오디오 품질 개선

MPAI 표준 기술 분야의 하나인 맥락 기반의 오디오 품질 개선 분야 (MPAI-CAE, MPAI Context-based Audio Enhancement)는 오디오 응용 분야에서 활용할 수 있는 큰 가능성을 보여주고 있다. 이 기술은 맥락 향상 정보를 입력된 오디오 콘텐츠에 적용해 가장 적절한 프로토콜로 오디오 출력을 전달하는 방식이다. 이 분야에서는 4개의 대표적인 유즈케이스가 제안됐는데, 오디오 녹음 보존(ARP, Audio Recording Preservation) 유즈케이스는 오디오 문서의 보존과 복구 과정을 다루는 것으로 역사적으로 중요한 오디오 문서를 디지털화한 복사본으로 변환해 보존하는 것이다. 오디오 전달(AOG, Audio-on-the-go) 유즈케이스는 사용자들이 서로 다른 환경하에서 상황별 안전과 청취의 품질을 개선할 수 있는 기술을 다루고 있다. 마지막으로 감성 개선 음성(EES, Emotion-Enhanced Speech) 유즈케이스에서는 다양한 수준의 감정을 포함한 음성을 생성하는 사용자 친화적인 시스템 제어 인터페이스를 구현하고 있다. 제안된 기술 중에 ARP 유즈케이스가 [그림 3]에 제시돼 있는데 여기에는 오래된 아날로그 테이프를 보존하기 위해 사용하는 접근방식과 다양한 매체에 담겨있는 음성 정보의 복구를



[그림 3] MPAI-CAE ARP 워크플로우[2]

위한 해결 방식이 기술로 정의돼 있다.

기술공모를 마치고 현재 표준기술 개발을 시작한 MPAI-CAE는 제안된 기술의 분석을 진행하고 있다. 제안된 기술 중의 하나는 오디오 테이프에 나타나는 다양한 불규칙성이나 이형태의 정의와 분류를 제안하는데[3], [그림 3]에서 구성하고 있는 오디오 개선 모듈과 분류 모듈 등 다양한 AI 모듈이 이러한 이형태를 인식해 분류하는 역할을 한다. 또한 음성 단락의 훼손을 복구하는 기술도 제안되었는데, 손상된 음성을 파악해 노이즈를 제거하고 AI 기술을 적용해 누락된 음성 부분을 복구하는 모듈도 시연했다. 이러한 기술들은 표준으로 정립돼 다양한 오디오 관련 분야에서 적용될 수 있으며 역사적인 자료의 복원에 사용될 수 있을 것으로 전망된다.

## 2.3 MPAI-MMC(MULTIMODAL CONVERSATION):

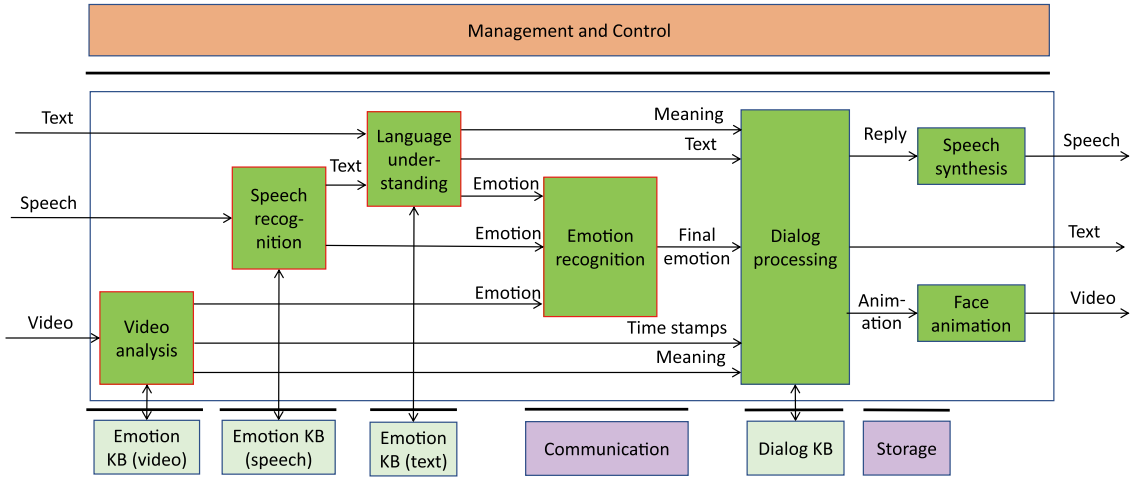
### 멀티모달 대화 응용 분야

MPAI에서 개발 중인 멀티모달 대화 표준은 사람과 사람 사이의 대화를 모사하는 방식으로 인공지능 기술을 사용해 사람과 기계 간의 대화가 가능하도록 하는 것을 목표로 한다. 현

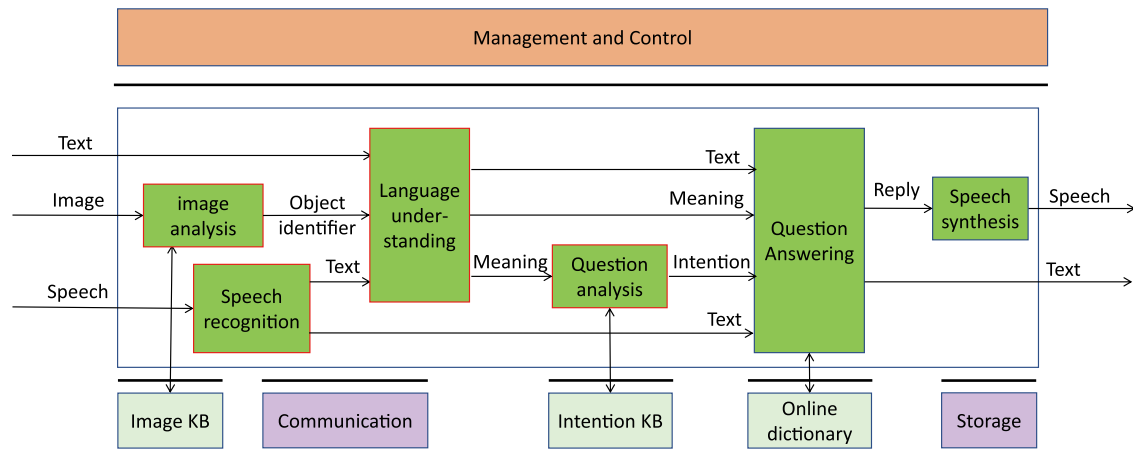
재, MPAI-MMC 표준은 다음과 같은 3개의 유즈케이스를 포함하고 있다. 감성 대화(CWE, Conversation With Emotion)는 [그림 4a]에 나와 있듯이, 사람이 하는 대화가 음성, 비디오, 텍스트로 입력돼 기계에 전달되며, 이에 대해 기계는 합성음과 생성얼굴을 출력으로 하여 대화에 응답하게 된다. 즉 기계는 인공지능 기술을 사용해 사람이 하는 말을 이해하고 사람의 말이나 표정에 나타나는 감정을 분석하여 적절한 응답을 찾아내는 방식으로 대화를 한다.

현재 MMC는 기술공모가 완료돼서 한국의 ETRI와 클레온이 감성대화 유즈케이스에 표준 기술을 제안했다. 주요 제안 사항으로는 각 AI 모듈에 대한 출력 포맷을 규정하는 기고서를 제출했다. 또한 언어 이해 AI 모듈의 감정 요소를 비롯해 얼굴생성 모듈 등 주요 요소 목록에 관련된 제안이 이뤄졌으며 기술공모의 한 과정으로 현재 평가 과정을 진행하는 중에 있다.

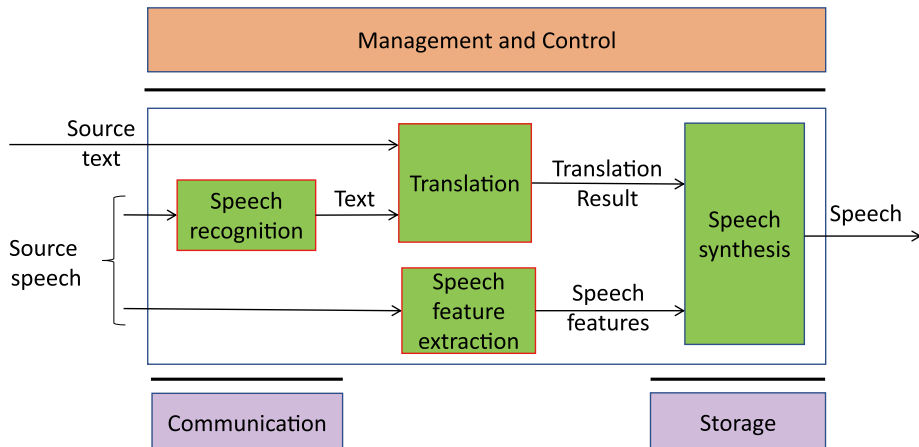
[그림 4b]에서 보여주고 있는 멀티모달 질의응답(MQA, Multimodal Question Answering) 유즈케이스는 화면에 보이는 물건에 대해 자연어로 질문을 하면 시스템이 합성 음성으로 거기



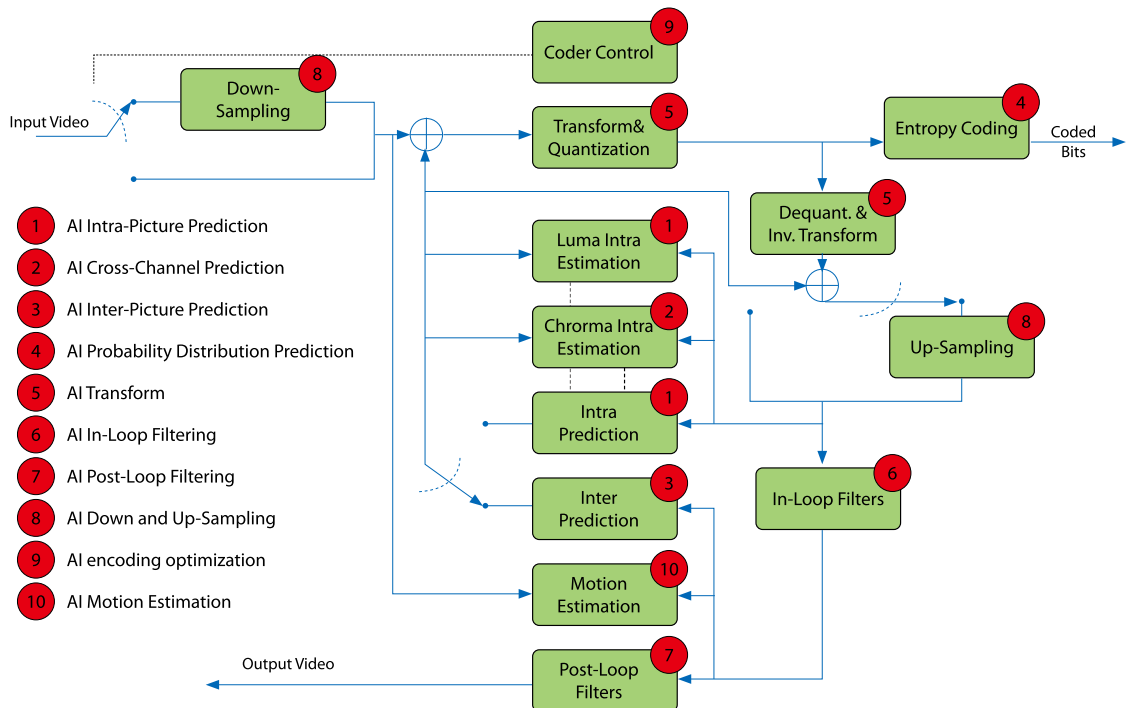
[그림 4(a)] MPAI-MMC: CWE[4]



[그림 4(b)] MPAI-MMC: MQA



[그림 4(c)] MPAI-MMC: PST



[그림 5] AI를 통한 전통적 비디오 코딩 방식의 개선[5]

에 대한 답을 제공하는 서비스에 대한 표준이다. 이 유즈케이스에 대해 제안된 기술에는 워크플로우상의 각 AIM에 대한 출력 포맷을 정형화해 규정하는 내용이 포함돼 있다. 또한 중요 구성 AIM인 질문분석 모듈의 질문 토픽, 포커스, 응답 타입, 도메인 요소 등에 대한 정의가 제공되어 질문의 의도를 표현하는 기술을 제안하고 있다. 질문분석 모듈에서 생성된 질문 의도는 언어 이해 AIM에서 산출된 질문의 의미와 함께 질의 응답 AIM에서 입력으로 사용돼 질문에 대한 적절한 응답을 생성하게 된다.

[그림 4c]의 개인화 자동통역(PST, Personalized automatic Speech Translation) 유즈케이스에서는 사람이 발성한 문장을 기계가 번역한 후에 합성 음성을 사용해 원 발화자의 음성 특성을 모방해 자연스러운 목소리로 출력하는 시스템을 보여준다. 기술공모에서 제안된 기술에

는 주요 구성 AI 모듈의 입력과 출력을 정형화해 정의하고 있다. 또한 이러한 자동통역 시스템을 구현하기 위해 필요한 사용자 인터페이스를 구체적으로 제시하고 상세한 서비스 방법을 시연했다. 이 시스템 방식은 클라우드 서비스를 활용해 무거운 시스템은 서버에 배치해 로컬 응용시스템의 부담을 낮추는 방식으로 빠른 속도를 구현했다. 또한 전통적인 통역서비스의 방식을 벗어나서 컴퓨터가 생성한 음성을 최대한 유저와 비슷한 특성을 갖도록 서비스를 구성하였다.

## 2.4 MPAI-EVC(EVC, Essential Video Coding) 비디오 코딩 응용분야

MPAI 비디오 그룹은 비디오 코딩 응용분야에 인공지능 기술을 적용해 높은 성능개선을 얻고자 다양한 실험을 수행하는 중이다. 논문 조사 결과로 인공지능 기술로 30% 정도의 성능개

선 효과가 가능하다는 사실을 알아냈다. 이를 적용하여 기존의 코딩 도구인 EVC에 AI 기술을 활용해 복잡성을 크게 높이지 않고도 EVC 코딩의 성능을 개선할 수 있는지 탐색하고 있다. [그림 5]는 참조 코덱 구성을 정의한다. 붉은 원은 인공지능 도구로 향상 또는 대체를 할 수 있는 데이터 처리 후보 블록을 나타낸다. 각각의 데이터 처리 도구를 위해서 세 가지의 단계가 필요한데 데이터 추출, 코드 블록의 훈련과 추론 단계가 여기에 해당된다. 각각의 코덱 블록에 대하여 EVC 코덱은 AI 프레임워크와 통신을 하고 하이브리드 EVC 코덱의 성능을 평가한다.

현재 MPAI는 위의 세 개의 단계를 구현하기에 필요한 구성 요소들을 개발하고 있다. 또한 최초의 AI 개선 데이터 처리 도구를 사용하기 위한 데이터를 수집하고 있으며 최종 개선 성능은 향후에 논문으로 발표될 예정이다.

### 3. 맺음말

최근에 창립된 사실 표준화 그룹인 MPAI는 오랜 기간 누적된 데이터 처리 기반의 디지털 미디어 압축기술을 기반으로 좀 더 넓은 범위의 데이터 코딩 분야로 적용 범위를 확산했다. 또한 오디오, 비디오, 자연어처리, 산업데이터, 게임 분야 등 다양한 분야의 유즈케이스와 관련된 표준을 개발하고 있다. MPAI의 접근 방법은 인공지능 기술의 사용에 중점을 두고 있지만 개별 인공지능 기술의 내용에 대해서는 블랙박스 접근법을 사용해 모듈의 재사용 방식을 표준화하고 있다. AI 시스템의 정의를 AI 프레임워크에서 시도되는 다양한 AI 모듈의 네트워크로 간주해 AI 시스템의 설명 가능성에 대한 동력을 제공한다. TTA

#### 참고문헌

- [1] MPAI Community (2021). Artificial Intelligence Framework. <https://mpai.community/standards/mpai-aif/#Technologies/> Last accessed 2021/05/03
- [2] MPAI Community (2021). Context-based Audio Enhancement. <https://mpai.community/standards/mpai-cae/> Last access: April 26, 2021.
- [3] Pretto, N., Fantozzi, C., Micheloni, E., Burini, V., and Canazza, S. (2019) Computing methodologies supporting the preservation of electroacoustic music from analog magnetic tape, Computer Music Journal, 42, no. 4, 59-74. [https://doi.org/10.1162/comj\\_a\\_00487](https://doi.org/10.1162/comj_a_00487).
- [4] MPAI Community (2021). Multimodal Conversation. <https://mpai.community/standards/mpai-mmc/> Last access: April 26, 2021.
- [5] Roberto Iacoviello, (2020). Analysis of performance of AI based video codecs. Input document to MPAI-EVC.