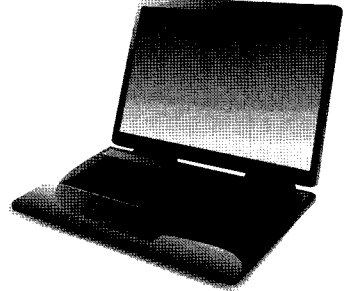


IPTV 자막 방송 표준



양진영 | TTA 데이터방송시험실 선임연구원
이은향 | TTA 데이터방송시험실 실장

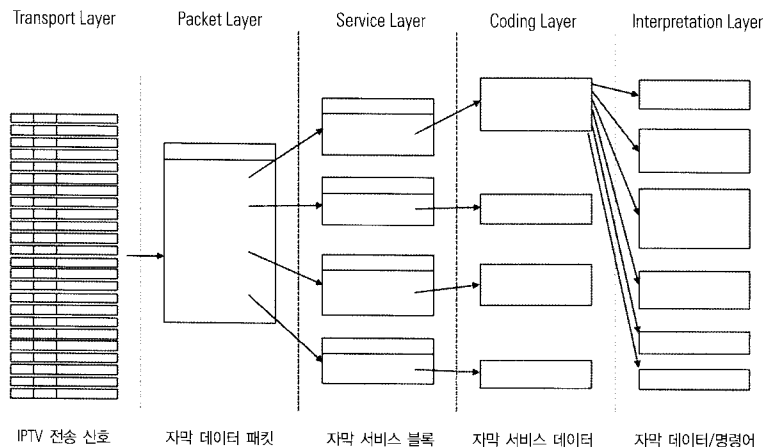
1. 머리말

IPTV 자막 방송 표준은 IPTV에서 자막 방송 서비스를 제공하기 위해 자막(closed caption) 데이터를 전송하고 수신하는데 요구되는 사항을 정의하고 있으며, 북미에서 자막 방송 규격으로 사용되는 CEA-708-D[1]를 기준으로 하고 있다. 이 표준은, 국내 IPTV 규격이 기본적으로 사용하는 비디오 압축 방식인 H.264[2]에 자막 데이터를 전송하는 방식으로 A/72[3]를 따르고 있고, 한국

어 자막을 송수신하기 위해 필요한 코드 세트와 전자/반자 처리 등에 대해 기술하고 있다. IPTV 자막 방송 표준은 2009년 9월 23일에 TTA 표준(표준번호: TTAK, KO-08,0022)으로 발행되었다.

2장에서는 자막 데이터를 전송하기 위한 프로토콜에 대해, 3장에서는 자막을 수신하기 위한 구현 요구사항 및 디코더 구조에 대해 소개하고자 한다.

2. IPTV 자막 방송 데이터의 전송



[그림 1] IPTV 자막 전송 프로토콜

2.1 프로토콜

IPTV 자막 방송을 위한 자막 데이터는 [그림 1]과 같이 전송 계층(transport layer), 패킷 계층(packet layer), 서비스 계층(service layer), 코딩 계층(coding layer), 해석 계층(interpretation layer) 등 5개 계층 구조에 따라 전송된다.

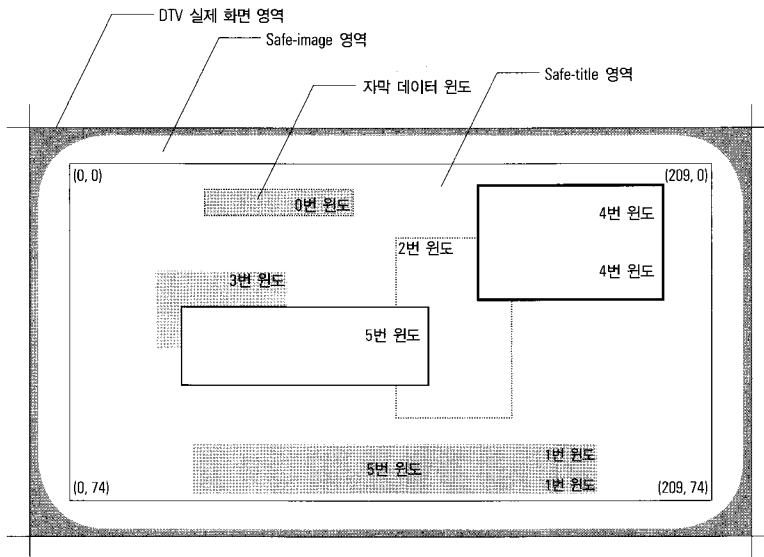
2.1.1 해석 계층

해석 계층은 [그림 2와 같이 자막 데이터의 TV 화면 상에서의 배치 정보가 어떤 형식으로 인코딩되고 디코딩되어야 하는지를 정의한다. 즉, TV 화면에서 자막 윈도의 크기 및 좌표 등의 배치 정보, 하나의 자막 윈도

내에 나타낼 수 있는 행과 열의 수를 정의한다. 또한 자막 문자가 표시될 때, 정렬 방법, 프린트 방향, 스크롤 방향, 컬러, 사이즈 등을 정의한다.

2.1.2 코딩 계층

코딩 계층은 자막 문자와 기호, 명령어에 코드값들을 할당한다. 영문 자막 데이터는 ANSI X3.41과 ISO 2022 표준에 따라, CL, GL, CR, GR의 4개 코드 그룹을 사용하여 코딩하고, 한글 자막 데이터는 <표 1>과 같이 KS X ISO/IEC 1046 한글 코드(유니코드)와 KS X 1001:2004 완성형 한글 코드를 사용한다.



[그림 2] 16:9 화면에서의 자막 윈도우 배치

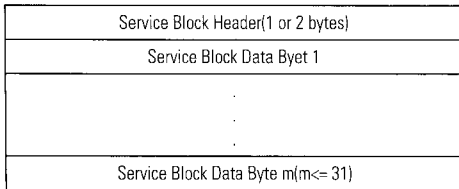
<표 1> IPTV에 적용되는 한글 자막 기본 문자표

구분	KS X ISO/IEC 10646(유니코드)		KS X 1001:2004 완성형 코드	
	블록 이름	범위(16진수)	블록 이름	범위(16진수)
영문 (로마자)	라틴(Basic Latin): 95자	0020-007E	1바이트 로마 문자(7bit): 95자	20-7E
	라틴 보충 1(Latin-1 Supplement): 96자	00A0-00FF		
한글	한글(Hangul): 11,172자	AC00-D7A3	2바이트 완성형 한글: 2,350자	B0A1-C8FE
특수문자 (약물)	KS X 1001:2004 완성형 코드의 2바이트 완성형 특수 문자와 동일 문자 집합: 986자		2바이트 완성형 특수문자: 986자	A1A1-ACFE
한자	7,744자		4,888자	

2.1.3 서비스 계층

자막은 서비스라고 불리는 여러 개의 논리적인 가상 채널들로 구성할 수 있다. 6개의 표준 서비스와 최대 57개의 확장 서비스를 추가할 수 있어서, 총 63개의 서비스를 제공할 수 있다. 6개의 표준 서비스(standard service) 중에서, 서비스 #1은 프로그램 오디오의 음성을 있는 그대로 텍스트로 옮긴 캡션을 담는 primary caption service로 할당되고, 서비스 #2는 primary caption service를 번역한 캡션을 담는 secondary language service로 할당된다.

각 서비스의 서비스 계층에는, 자막 데이터 채널 서비스 개수, 서비스 유형, 서비스 속성 등을 헤더에 정의한다. [그림 3]과 [그림 4]가 자막 서비스 블록의 구조와 구문 구조를 나타내고 있다.



[그림 3] 자막 서비스 블록

```

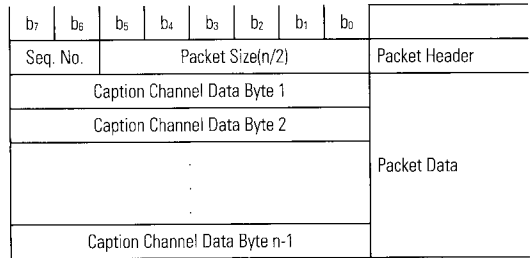
service_block {
    /* service block header */
    service_number      : 3 bits
    block_size          : 5 bits
    /* >=0 &&<3 */

    /* extended service block header extension */
    if(service_number == b'111 && block_size !=0)
    {
        null_fill
        : 2 bits
        extended_service_number
        : 6 bits
    }
    /* block data sequence */
    if (service_number !=0)
    {
        for(i=0; i<block_size; i++)
        {
            Block_data[i] : 8 bits
        }
    }
}
    
```

[그림 4] 자막 서비스 블록의 구문 구조

2.1.4 패킷 계층

서비스 계층에서 부호화된 각 캡션 서비스 데이터는 패킷 형태로 구성된다. [그림 5]와 [그림 6]이 캡션 채널 패킷의 구조와 구문을 나타내고 있다.



[그림 5] 캡션 채널 패킷의 구조

```

caption_channel_packet {
    /* packet header */
    sequence_number: 2bits
    packet_size: 6bits

    /*packet data sequence*/
    /*if(packet_size_code==0)*/
    /*packet_data_size=127*/
    /*else*/
    /*packet_data_size=(packet_size_code*2)-1*/
    for(i=0, i<packet_data_size; i++)
    {
        Packet_data[i] : 8bits;
    }
}
    
```

[그림 6] 캡션 채널 패킷의 구문

2.1.5 전송 계층

패킷 계층의 데이터는 H.264의 video elementary stream에 실어서 전송된다. A/72의 6.4절에 따라, SEI(Supplemental Enhancement Information) RBSP(Raw Byte Sequence Payload)의 payload에 실어

<표 2> SEI RBSP payload의 구성

Field Name	Length	Alignment
user_data_registered_itu_t_t35 () {		
itu_t_t35_country_code	8	bslbf
itu_t_t35_provider_code	16	bslbf
user_identifier	32	bslbf
user_structure ()		
}		

〈표 3〉 user_structure()의 구성

구문	비트수	형식
user_structure() {		
process_em_data_flag	1	bslbf
process_cc_data_flag	1	bslbf
additional_data_flag	1	bslbf
cc_count	5	uimsbf
em_data	8	bslbf
for (i=0;i<cc_count;i++) {		
marker_bits	5	'1111 1'
cc_valid	1	bslbf
cc_type	2	bslbf
cc_data_1	8	bslbf
cc_data_2	8	bslbf
}		
marker_bits	8	'1111 1111'
if(additional_data_flag) {		
while (nextbits()!='0000 0000 0000 0000 0000 0001') {		
additional_user_data	8	
}		
}		
}		

서 전송한다. 〈표 2〉가 SEI Rbsp payload의 구문을 보여 주고 있으며, 그 중에서 user_structure()의 구문 구조는 〈표 3〉과 같다.

2.2 시그널링

자막 서비스가 전송될 경우, 자막 서비스가 전송됨을 알리는 자막 서비스 서술자(Caption Service Descriptor)를 MPEG-2 TS(Transport Stream)의 PMT(Program Map Table)에 실어서 전송한다. 자막 서비스 서술자에는 자막 서비스의 개수, 각 자막 서비스에 대한 언어 코드가 실리며, 또한 자막 언어가 한글일 경우 유니코드인지 완성형 한글 코드인지를 나타내는 정보가 실린다. 〈표 4〉가 자막 서비스 서술자의 구문 구조이다.

〈표 4〉 자막 서비스 서술자의 구문 구조

구문	비트수	형식
caption_service_descriptor() {		
descriptor_tag	8	0x86
descriptor_length	8	uimsbf
reserved	3	'111'
number_of_services	5	uimsbf
for (i=0;i<number_of_services;i++) {		
language	8*3	uimsbf
cc_type	1	bslbf
reserved	1	'1'
if (cc_type==line21) {		
reserved	5	'11111'
line21_field	1	bslbf
}		
else		
caption_service_number	6	uimsbf
easy_reader	1	bslbf
wide_aspect_ratio	1	bslbf
korean_code	1	bslbf
reserved	13	'11111111111111'
}		
}		
}		

3. IPTV 자막 방송의 수신

3.1 디코더 필수 구현 사항

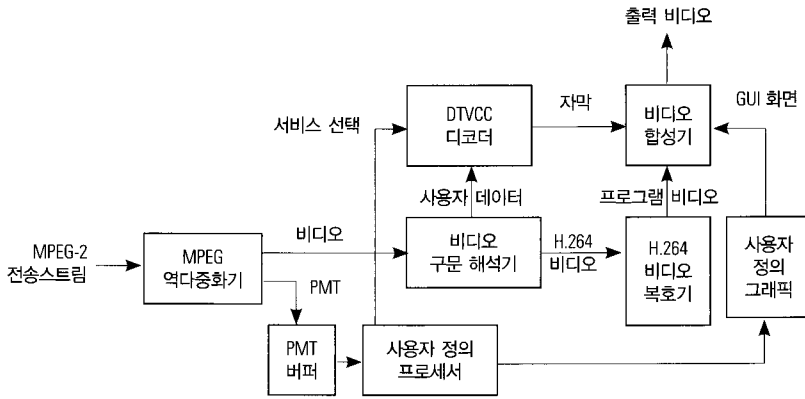
IPTV 자막 방송을 위해 디코더가 제공해야 하는 최소 요구사항은 다음과 같다.

- 대역폭: 어느 1초 구간을 평균하더라도 하나의 서비스가 전체 캡션 대역폭 9,600bps의 25%를 초과할 수 없음. 따라서, 디코더는 서비스당 최대 2,400bps의 캡션 데이터를 처리할 수 있는 성능과 버퍼를 구비해야 함
- 자막 서비스 개수: 6개의 표준 서비스는 반드시 지원할 수 있어야 함

- 자막 원도 해상도: 가로는 16:9인 경우 전자는 26자이고 반자는 52자이며, 4:3인 경우 전자는 20자이고 반자는 40자를 기준으로 정함. 세로는 12줄을 기준으로 정함
- 자막 원도 개수: 한 번에 적어도 4개의 caption window를 디스플레이할 수 있어야 함. 디코더는 최소한 8개 row의 캡션을 지원할 수 있도록 저장소를 유지하고 있어야 함(4개 row의 캡션 원도를 디스플레이하고 디스플레이되지는 않지만 다음 4개 row의 캡션을 버퍼링하고 있는 원도를 지원하는 최악의 경우를 지원하기 위함)
- 자막 원도 행과 열 고정: 디코더는 원도의 행과 열의 개수가 고정되어 있다고 가정해야 함. 따라서 사용자가 캐릭터 크기를 크거나 작게 조정하더라도 워드 래핑(word wrapping)이나 언래핑(un-wrapping)이 일어나지 않음
- 정렬: 좌측정렬, 우측정렬, 중앙정렬은 반드시 지원해야 함
- 프린트 방향: 캐릭터를 화면에 쓸 때의 방향은 LEFT_TO_RIGHT를 반드시 지원해야 함
- 스크롤 방향: 캐리지 리턴(carriage return)을 만났을 때, 텍스트를 스크롤하는 방향은 BOTTOM_TO_TOP을 반드시 지원해야 함
- 디스플레이 효과: 보이지 않는 원도와 보이는 원도를 사용하여 캡션이 즉시 표시(pop-on)되는 SNAP display effect를 기본으로 제공해야 함
- 원도 컬러와 경계(border): 경계가 없는 불투명 검정색(solid black) 바탕의 원도와 경계가 없는 투명 바탕의 원도는 반드시 지원해야 함
- predefined window / predefined pen style: Predefined window style 1과 predefined window style 2는 반드시 지원해야 함. 그 외의 수신기가 지원하지 않는 predefined window style은 predefined window style 2로 간주함. 또한 predefined pen attribute style 1을 지원해야 함
- 펜 크기: small, standard, large 사이즈를 지원해야 함
- 펜 스타일: normal, italic, underline을 지원해야 함
- 폰트 스타일: Korea iTV SansSerifD를 지원해야 함
- foreground color/opacity: 불투명 및 플래시(flash)를 반드시 지원해야 하며, 기본 8가지 컬러(white, black, red, green, blue, yellow, magenta, cyan)를 지원해야 함
- Background color/opacity: 투명, 불투명, 반투명은 반드시 지원해야 하며, 기본 8가지 컬러를 지원함
- 컬러 표현: 기본 8가지 컬러(black, white, red, green, blue, yellow, magenta, cyan: 각각에 대한 RGB값 정의)는 반드시 지원해야 함. 규격에 정의되지 않은 RGB값의 컬러를 수신할 경우, CEA-708-D에 정의된 알고리즘에 따라 컬러를 매핑함
- 자막 화면 자동 clear 기능: CEA-708-D 혹은 IPTV 자막 방송 표준을 따르는 자막 데이터를 받아 화면에 자막 정보를 표시한 후 16초 이상 다음 자막 데이터가 전송되지 않을 경우 수신기에서는 표시된 자막 화면을 자동으로 Clear 함

3.2 디코더 구조

IPTV 수신기의 자막 데이터 수신부의 블록도는 [그림 7]과 같이 구성할 수 있다. MPEG-2 TS 스트림은 MPEG 역다중화기에 의해 비디오 ES가 필터링되어 비디오 구문 해석기로 전달되면, 자막 데이터가 실린 사용자 데이터는 DTVCC 디코더로 전달되고, 나머지 H.264 압축 비디오 데이터는 H.264 복호기에서 디코딩된다. DTVCC 디코더에서 복호화된 자막 데이터는, 사용자 정의 그래픽에서 사용자가 자막을 on 시키면, 비디오와 함께 비디오 합성기에서 overlay되어 출력된다. 한편, 자막 서비스 서술자는 PMT 버퍼에서 사용자 정



[그림 7] IPTV 수신기의 자막 데이터 수신부

의 프로세서로 전달되어 사용자가 선택하는 서비스가 무엇인지를 알려주어 DTVCC 디코더가 해당 서비스 패킷을 복호화하게 한다.

4. 맺음말

자막은 TV 프로그램을 구성하는 오디오를 시각적으로 전달하는 데이터로서, 비디오에 겹쳐서 제공되며 사용자의 요구에 따라 선택적으로 서비스될 수 있는 기술이다. 또한, 하나의 A/V 서비스에는 오디오를 있는 그대로 텍스트로 옮긴 자막, 좀 더 쉽게 이해할 수 있게 표현한 자막, 다른 여러 가지 언어로 표현한 자막 등 여러 종류의 자막 서비스 제공이 가능하다. 따라서 자막

방송은 청각 장애인의 TV 프로그램 이해에 도움이 될 뿐만 아니라, 내국인의 외국어 교육, 다양한 언어를 사용하는 외국인들의 한국어 프로그램 이해, 해외 제작 영상물의 이해 등 여러 분야에 적용될 수 있을 것이다.

[참고문헌]

- [1] CEA-708-D, 'Digital Television^(DTV) Closed Captioning', August 2008.
- [2] H.264, 'Advanced video coding for generic audiovisual services,' March 2009.
- [3] A/72, 'Video System Characteristics of AVC in the ATSC Digital Television System,' July 2008.
- [4] CEA-608-E, 'Line 21 Data Services,' April 2008. **TTA**