

## 스마트링크 TV: 콘텐츠 장면기반 2<sup>nd</sup> 스크린 서비스 플랫폼

최대훈, 이재호, 박성춘, 이동준  
한국방송공사 기술연구소  
{maxhoon, jaeho, sparko, djlee}@kbs.co.kr

### SmartLink TV: Content Scene-based 2<sup>nd</sup> Screen Service Platform

Daehoon Choi, Jaeho Lee, Sungchoon Park, Dongjune Lee  
Technical Research Institute, KBS

#### 요 약

오디오 핑거프린트는 콘텐츠의 오디오 신호의 특색을 요약한 서명(signature)으로 2nd 스크린 서비스의 기반 기술로 주목 받고 있다. 이를 이용한 오디오 핑거프린트 시스템은 콘텐츠의 오디오를 인식하고 해당 콘텐츠와 연관된 메타 데이터를 2nd 디바이스와 웹을 통해 연결시키는 기능을 제공한다. 스마트링크 TV 는 이러한 오디오 핑거프린트를 이용한 KBS 의 콘텐츠 장면(scene) 기반 2nd 스크린 서비스 플랫폼이다. 본 논문에서는 한국의 방송환경에 적용된 스마트링크 TV 의 전반적인 플랫폼 구조와 활용방법에 대해 살펴본다.

#### 1. 서론

최근 모바일 스마트 디바이스 보급률의 증가와 함께 다양한 웹기반 미디어 플랫폼들이 등장함에 따라, TV 시청행태에 있어 주목할만한 변화가 있다. 과거 대표적인 린백(Lean-back) 미디어인 TV 는 가족의 공용 디바이스로, 영상 시청에 최적화되어 높은 신뢰도가 장점인 반면 영상 시청 이외의 기능들에 대해서는 대부분 대체제가 존재하고 개인화하기 어려운 단점이 남아 있었다.

모바일 스마트 디바이스의 등장은 이러한 TV 의 단점을 보완하는 보완재의 역할을 수행하기에 적합한 조건을 갖추기 시작한 계기가 되었다. 개인형 디바이스로 정보 검색에 최적화된 모바일 스마트 디바이스는 편리한 조작성과 휴대성을 바탕으로 TV 의 동반 디바이스로서(companion device)의 역할을 수행하기 적합한 환경을 갖추게 되었다. 이에 따라, TV 시청을 방해하지 않으면서 TV 시청 도중에 해당 콘텐츠와 연관된 부가 정보를 웹기반의 2nd 스크린으로 소비하는 시청 행태가 뚜렷이 나타나고 있다[1]. 또한, TV 시청 도중 2nd 디바이스로 멀티태스킹하는 시청패턴과 더불어 소셜 미디어 상에 현재 시청중인 TV 프로그램 관련 포스팅이 증가하는 등 TV 시청과 소셜 미디어를 동시에 이용하는 시청행태가 확산되고 있다[2]. 또한, TV 가 점차 2nd 디바이스를 이용한 린포워드(Lean-forward) 미디어로 진화하고 있으며 이에 따라 Social TV[3] 라 불리는 플랫폼 사업자가 등장하고 있다.

스마트링크 TV 는 TV 시청의 편의성과 스마트 단말의 편리한 웹 접근성을 결합한 2nd 스크린 서비스의 기반 플랫폼으로, 위에 언급된 시청행태 변화에 대응 하기 위한 기반 기술로 오디오 핑거프린트(Audio Fingerprint) 기술을 활용하고 있다. 실제 오디오 핑거프린트 기술은 과거에 방송 콘텐츠 모니터링, 음악식별 및 저작권 필터링의 용도로

활용되었다. 그러나 Gracenote[4], Shazam[5], IntoNow[6], SOTY[7] 와 같은 최근의 오디오 핑거프린트 시스템에서는 방송 신호에 대한 일체의 데이터를 삽입을 배제한 채 시청중인 콘텐츠의 오디오를 인식하고 해당 콘텐츠와 연관된 메타 데이터를 2nd 디바이스와 동기화시키는 기술로 활용하고 있다. 이들 콘텐츠 기반 서비스들은 시청자로 하여금 웹을 통해 보다 편리한 방식으로 2<sup>nd</sup> 스크린 서비스로의 접근성을 높이기 위한 용도로 활용되고 있다.

스마트링크 TV 는 이러한 콘텐츠 기반 서비스에서 한단계 진화된 KBS 의 콘텐츠 ‘장면(scene) 기반’ 2nd 스크린 서비스 플랫폼이다. 스마트링크 TV 플랫폼은 최신의 오디오 핑거프린트 기술을 방송환경에 적용하여 ‘2nd 디바이스와의 연결’ 뿐 아니라, 방송사업자가 소유한 방대한 콘텐츠와 해당 콘텐츠의 장면(scene) 기반의 메타데이터를 활용한다. 이를 통해 시청자의 시청 장면에 맞는 방송정보, 뉴스, 장소, PPL 광고, 음악, SNS 등의 정보를 제공하며 VOD 콘텐츠 뿐만 아니라 온에어(OnAir) 방송 콘텐츠에 대해서도 대응이 가능한 특징이 있다. 본 논문에서는 이러한 방송사업자의 오디오 핑거프린트 기술에 기반한 2<sup>nd</sup> 스크린 서비스 플랫폼으로서의 스마트링크 TV 의 구조와 활용방법에 대해 소개한다.

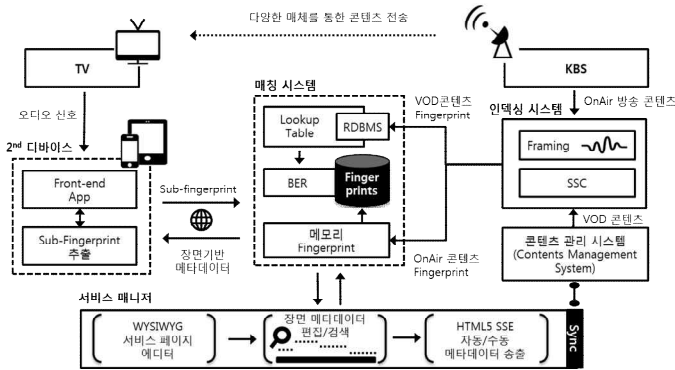
#### 2. 플랫폼 구조 및 구현

[그림 1]은 스마트링크 TV 플랫폼의 구성도로 웹 기반의 2nd Screen 서비스를 위한 하위 시스템의 구조를 포함한다. 스마트링크 TV 는 다음과 같은 하위 시스템들로 구성되어 있다

##### 2.1 인덱싱 시스템

스마트링크 TV 인덱싱 시스템은 자동콘텐츠인식(ACR) 을

위한 기반 기술로서 오디오 핑거프린팅 방식을 채택하였다. 오디오를 이용한 대표적인 자동인식기술로는 오디오 핑거프린팅과 워터마크 방식이 존재하지만 [표 1]과 같이 콘텐츠 제공자인 지상파 방송사업자 관점에서는 핑거프린팅 방식을 선호하는 추세이다.



[그림 1] 스마트링크 TV 플랫폼 구성도

[표 1] 오디오 핑거프린팅과 워터마크 기술 비교

구분	핑거프린팅 기술	워터마크 기술
장점	- 기배포 미디어 인식가능 - 콘텐츠 변형의 부담 없음	- 동일한 오디오 구별 가능 - 시스템 구축 투자비용 적음
단점	- 동일(유사)한 오디오 구별 불가 - 시스템 구축 투자 비용 높음	- 미디어 배포전 워터마크 삽입이 필수 - 워터마크 삽입에 따른 콘텐츠 변형 필요
특징	- 콘텐츠 제공자 서비스 및 비즈니스 모델 적용 유리 - 방송 인식을 위해 probe mode 로 시스템 연동하기 때문에 적용성이 높음	- 플랫폼 사업자 서비스 및 비즈니스 모델 적용 용이 - 방송 인식을 위해 Inline mode로 시스템에 삽입되기 때문에 장에 관리에 민감함

따라서, 인덱싱 시스템은 온에어로 현재 방송중인 콘텐츠 및 VOD 콘텐츠의 오디오 신호를 받아 실시간으로 오디오 핑거프린트를 추출한다. 핑거프린트 추출과 관련된 연구분야에는 다양한 알고리즘이 존재한다.

대표적인 것이 Haitsma의 알고리즘[8]으로 각 370ms 오디오 프레임의 에너지를 구하고 서브 밴드별 에너지의 차를 구한 후, 직전 프레임과의 차를 다시 구해서 최종적으로 이 값을 핑거프린트로 사용한다. 본 논문에서 사용된 핑거프린트 알고리즘은 음성인식(speech recognition)에서 많이 사용된 SSC(Spectral Subband Centroid)[9]를 기반으로 하고 있으며, 이를 이용하여 원본 콘텐츠의 오디오 핑거프린트를 추출하고 해당 콘텐츠의 핑거프린트를 검색을 위한 데이터베이스로 구축한다.

## 2.2 매칭 시스템

매칭 시스템은 사용단말로부터 추출된 오디오 핑거프린트에 대한 콘텐츠 인식 결과를 제공하는 시스템이다. 일반적으로 핑거프린트 매칭의 목적은 검색요청(query)이

들어왔을 때, 특정 자료구조(또는 인덱스)를 생성하여 오디오 핑거프린트 데이터베이스 검색 수행(operation)의 수를 줄이는 것이다. 따라서, 비교 후보군을 줄이고 오버헤드가 많이 드는 단순비교(Brute-force) 방식을 피하기 위해 인덱싱을 사용하는 아이디어가 CBID(Content based Identification) 분야에서 연구되어왔다[10].

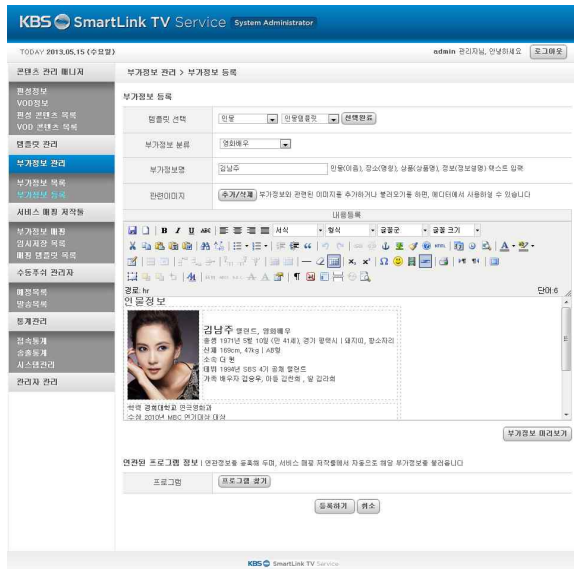
[8]에서 Haitsma는 핑거프린트 인덱싱을 위한 두 단계의 해쉬 방법론(Hashing scheme)을 제안하였다. 콘텐츠의 위치를 가리키는 해쉬 인덱스를 사용하여 전체 오디오 핑거프린트 중 일부본인 서브 핑거프린트(sub-fingerprint)의 정확한(exact) 매칭을 제공함으로써, BER(Bit Error Rate) 계산의 최소화를 위한 후보 콘텐츠의 리스트와 위치를 효율적으로 생성하였다. 매칭 시스템에서는 Haitsma의 알고리즘과 유사하나 특성이 다른 VOD와 온에어 방송용 콘텐츠 대응을 위해 상이한 접근 방식을 취한다.

대용량 오디오 핑거프린트로 구성된 VOD 콘텐츠일 경우 [11]에서처럼, 시스템의 안정성과 검색의 효율성을 확보하기 위해 룩업테이블(Lookup table-해쉬 구조의 첫 단계)을 메인 메모리 대신 SSD(solid state drives) 이차 저장장치 기반의 RDBMS(Relational Database Management System)로 구축하였다. 이와 달리, 많은 네트워크 트래픽을 유발시키고 실시간 응답을 강조하는 온에어 방송 콘텐츠에 대해서는 룩업테이블 구축 단계를 생략한다. 즉, 룩업테이블 구성을 위한 오버헤드(overhead)를 줄이고 서브 핑거프린트와 핑거프린트(실시간으로 추출되는 2~5분 가량의 오디오 핑거프린트)의 매칭을 위한 BER 계산을 곧장 메인 메모리에서 진행함으로써, 실시간 응답을 가능하게 한다.

메모리에 존재하는 오디오 핑거프린트는 VOD 콘텐츠로의 활용을 위해 콘텐츠 단위로 핑거프린트 데이터 베이스에 저장된다. 이후 이원화된 매칭을 통해 온에어 및 VOD 콘텐츠에 대한 동시 검색이 가능하다. 결과적으로, 스마트링크 TV 플랫폼은 인덱싱 시스템과 함께 오디오 핑거프린트 시스템이 가져야 할 세가지 특성 [12]과 더불어 안정성을 확보하였다.



[그림 2] 장면기반 메타데이터 매핑



[그림 3] 서비스 페이지 편집기능

### 2.3 서비스 매니저 시스템

서비스 매니저 시스템에서는 [그림 2]와 같이 콘텐츠 타임라인(Timeline) 상에 존재하는 장면 기반의 메타데이터 [13]의 조회와 편집이 가능하다. 메타데이터 편집은 콘텐츠의 특정 재생구간에 해당 재생구간과 연관된 방송 부가정보를 매핑하는 것을 말한다. 또한, 방송사업자의 CMS(Content Management System)와 연동하여 방송편성정보를 기반으로 각 콘텐츠의 메타 데이터를 관리하게 된다. 매칭 시스템으로부터 사용자가 인식 요청을 한 콘텐츠의 ID와 시청 시점정보가 얻어지면, 서비스 매니저는 해당 콘텐츠 장면의 시점정보에 상응하는 메타 데이터를 콘텐츠 인식 요청의 결과로 전달하게 된다.

기술적으로, W3C HTML5 SSE(Server Sent Event)표준 스펙 [14]을 만족하는 웹(Web) 기반 서비스 매니저는 최초에 인식된 시청자의 콘텐츠 시청시점부터 장면이 바뀌어감에 따라(인식 시점부터 시간이 지남에 따라) 장면기반의 메타데이터를 시청자에게 자동으로 SSE를 통해 전달(PUSH)한다. 또한, 관라자의 필요에 따라 특정 콘텐츠를 시청하고 있는 시청자(인식이 완료되어 서비스 매니저와 세션을 유지하고 있는)에게 수동으로 즉시 메타데이터를 전달할 수 있다. 시청 시점은 주기적으로 업데이트되며 장면기반 메타데이터 전송은 VOD 콘텐츠뿐만 아니라 온에어 콘텐츠에도 대응이 가능하다.

스마트링크 TV의 Web 기반 프론트 엔드(front-end) 애플리케이션이 전달받은 메타데이터는 결과적으로 웹페이지의 URL이다. 애플리케이션은 콘텐츠의 인식결과로 전송받은 URL로 접속하여 웹페이지를 사용자에게 보여줌으로써 방송 프로그램과 관련된 부가 정보를 서비스 받게 된다. 따라서, 서비스 매니저 시스템에서는 다양한 메타데이터 즉, HTML 웹 페이지를 손쉽게 저작 및 편집하는 툴(tool)이 필요하게 된다. 이러한 필요성으로 인해 서비스 매니저 시스템은 웹 기반의 서비스 페이지 저작툴로서 [그림 3]과 같이 WYSIWYG [15] 에디터 기능을 포함하게 되었다.

결과적으로, 이러한 서비스 매니저 시스템의 주요 기능들을

통해 스마트링크 TV 플랫폼은 장면 기반의 다양한 2nd 스크린 응용 서비스에 적용할 수 있다.

### 3. 결론

본 논문을 통해서 KBS의 2<sup>nd</sup> 스크린 서비스를 위한 스마트링크 TV 플랫폼의 구조와 활용 방법에 대해 살펴보았다. 스마트링크 TV는 (i) 프론트 엔드 애플리케이션을 이용하여 온에어 콘텐츠와 VOD 콘텐츠의 장면과 연관된 부가 정보를 보여준다. 이를 위해 (ii) 핑거프린트 인덱싱 시스템은 온에어 및 VOD 콘텐츠의 오디오 핑거프린트를 실시간으로 생성하여 이를 데이터베이스로 구축하며 (iii) 매칭 시스템에서는 수많은 시청자들의 온라인 콘텐츠 인식을 대응한다. (iv) 서비스 매니저 시스템은 사용자의 콘텐츠 인식의 결과로 장면기반의 메타데이터의 관리 및 자동/수동 전송뿐 아니라 효과적인 서비스 페이지 저작 기능도 제공한다.

### 참고문헌

- [1] Nielsen, "State of the media: Cross-platform report", Q2 2012
- [2] Ericson, "An Ericsson Consumer Insight Summary Report", 2011
- [3] Pablo Cesar, David Geerts, "Past, present, and future of social TV", in *IEEE International workshop on Social TV*, 2011
- [4] Gracenote. <http://www.gracenote.com/>.
- [5] Shazam. <http://www.shazam.com/>.
- [6] Intonow. <http://www.intonow.com/>.
- [7] SOTY. <http://www.sotyweb.sbs.co.kr>
- [8] J.Haitsma and T.Kalker, "A Highly Robust Audio fingerprinting System", in *International Conf. on Music information Retrieval*, 2002.
- [9] Wei Xiong, "Audio fingerprinting based on dynamic subband locating and normalized SSC", in *International Conf. on Audio, Language and Image Processing*, 2012.
- [10] Pedro Cano, Eloi Batlle, T.Kalker, and J.Haitsma, "A review of algorithms for audio fingerprinting", in *Proceedings of the IEEE International workshop on Multimedia Signal Proc-essing*, pages 169-173, 2002
- [11] Hendrik Schreiber, Peter Grosche and Meinard Muller, "A re-ordering strategy for accelerating index-based audio fingerprinting", in *International Conf. on Music information Retrieval*, 2011
- [12] Jin S. Seo, Minho Jin, Sunil Lee, Dalwon Jang, Seunjae Lee, and Chang D. Yoo, "Audio fingerprinting based on normalized spectral subband moments", *IEEE Signal Processing Letters*, vol. 13, pp.209-212, 2006.
- [13] 디지털편성위원회, "KBS 방송메타데이터표준 v1.2", 2011
- [14] W3C Working Group. <http://www.w3.org/TR/eventsource/>
- [15] <http://en.wikipedia.org/wiki/WYSIWYG>