

# 가라오케의 시간동기화를 위한 효율적인 시간정보의 표현모델

황준식, 최무락, 김성민, 고성제  
고려대학교 전자컴퓨터공학과

e-mail : {jshwang, mrchoi, smkim and sjko}@dali.korea.ac.kr

## Effective Representation of Time Information for Synchronization in Karaoke

Jun-Sik Hwang, Moo-Rak Choi, Sung-Min Kim and Sung-Jea Ko  
Department of Electronics & Computer Engineering  
Korea University

### Abstract

As different media formats converge into one digital content, we need to put time information in the content to synchronize those media. In this paper, we propose more efficient methods to represent time information aimed at Karaoke application. We utilize the linguistic and musical characteristic of Karaoke. Especially we make a novel codeword rule for the beat of music. In our experiment, our proposed methods turn out to drastically reduce time information.

### I. 서론

서로 다른 종류의 멀티미디어 콘텐츠를 융합하기 위해서는 그에 대한 시간동기화 정보가 필요하다. 특히 가라오케의 경우 오디오를 주축으로 하여 텍스트 및 비디오 정보가 결합된 복합적인 콘텐츠이다. 가사 텍스트를 음원 반주에 맞추어 한 줄씩 디스플레이 하기 위해 동기화를 위한 문법을 정의한 후에 소프트웨어 어플리케이션으로 구현한 경우도 있다[1].

본 논문에서는 가라오케에서 콘텐츠간의 시간동기화를 위해 정보량 측면에서 효율적인 시간정보의 표현 모델을 제안한다.

### II. 기존 동기화 방법

가라오케의 경우 반주에 따른 가사의 적절한 디스플레이를 위해 서로 다른 이벤트에 따라 marker type을 정의하고 시간정보를 부여한다. marker type은 “노래 시작”, “반주 시작”, “coloring”, “가사 줄바뀜” 등 여러 종류가 있다. 가사 컬러링의 경우 그림 1과 같이 하나의 문자를 몇 개의 segment로 균등하게 나누어 각 segment에 대한 색변환 시점을 정의하고 있다. 음원 재생기는 음악을 연주하면서 재생시간과 각 marker의 시각을 비교해 주어진 event를 수행하게 된다[2].

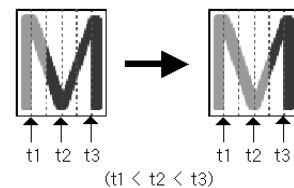


그림 1. 문자 컬러링 과정

가사와 음악데이터의 동기화를 위한 시간정보는 경과된 시간으로 표현되는데, 예를 들자면 각 이벤트의 발생 시각을 시, 분, 초와 msec 단위(00:00:00:000)로 나타내며 각각의 시각에는 고정된 바이트가 할당된다[3]. 이러한 기존의 방법은 직관적이긴 하지만 불필요한 정보의 중복성이 존재한다.

### III. 제안하는 방법

본 논문에서는 정보량측면에서 기존의 방법보다 효율적인 2가지 방안을 제안한다. 첫 번째는 언어적 특성을 이용하여 불필요한 시간정보를 줄이는 것이며 두 번째는 음악적 특성을 고려하여 고정된 바이트 대신 가변적인 비트열로 시간정보를 표현하는 것이다.

첫 번째 방법은 문자가 아닌 음절 단위로 시간정보를 부여하는 것이다. 한글의 경우 하나의 문자로 하나의 음절을 나타내지만 영어의 경우 하나 혹은 몇 개의 문자가 조합되어 한 음절을 표현한다. 그림 2의 (b)와 같이 음절에 따라 시간정보를 할당하는 것이 각각의 문자에 할당하는 (a)보다 효율적임을 보여준다.

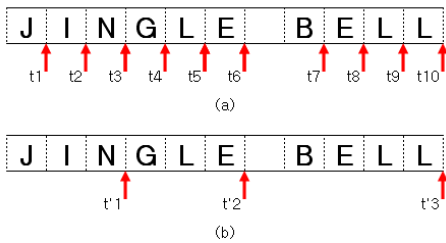


그림 2. 시간정보 정렬방식의 비교

두 번째는 모든 시간정보를 기준시간에 대한 상대적인 ratio로 표현하는 것이다. 예를 들어 기준시간을 1000 msec로 설정할 경우 500 msec는 기준시간의 '1/2'로 나타내며 1500 msec는 '3/2' 등으로 표현된다. 그리고 이러한 ratio는 본 논문에서 제안하는 codeword 규칙에 따라 가변적인 bit string으로 변환한다.

생성 규칙은 다음과 같으며 bit에 따른 2가지 계산방식을 정의하고 있다. bit '1'은 선행비트로부터 전달받은 ratio를 2로 나누고, 나뉜 ratio를 우측으로 전달한다. bit '0'은 덧셈 역할을 하며 전달받은 ratio를 그대로 우측으로 전달한다. 그림 3은 bit string으로부터 ratio가 어떻게 생성되는지를 단계별로 도시하고 있다.

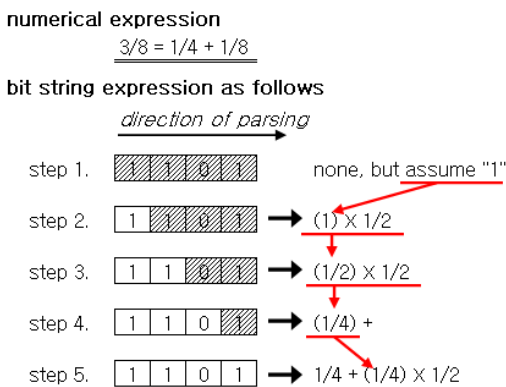


그림 3. bit string에서 ratio로 변환하는 예

표 1은 전술한 codeword 규칙에 따른 몇 가지 비트 열의 예를 나열하고 있다. 이 표를 살펴보면 기준시간 대비 factor 2 만큼 감소하거나 정수배로 증가하는 ratio의 경우 bit string이 상대적으로 짧다는 특징을 알 수 있다.

표 1. bit string 표현의 예

ratio	bit string	bit length
1/2	1	1
1/4	11	2
1/8	111	3
1	0	1
2	00	2
3	000	3
3/4	101	3
5/8	1011	4
33/16	001111	6

### IV. 실험 결과

본 실험에서는 영문 캐롤 곡인 “징글벨”을 통해 비교실험을 행하였다. 기존의 방식에서는 문자를 6개의 segment로 나눈 후 각 segment에 4byte의 고정된 코드 길이로 발생시각을 표현하였다. 제안하는 방식에서는 한 마디의 시간 길이를 기준시간으로 설정하였으며 가변되는 bit string의 구분을 위해 string 말미에 ‘100’ tag를 넣었다.

가사 컬러링에 대한 시간정보만을 가지고 실험한 결과로는 표 2와 같이 제안하는 시간 모델이 기존 모델에 비해 시간정보를 약 110배의 비율로 압축하였다.

표 2. “징글벨”에 대한 정보량 실험 결과

모델	시간정보량에 대한 ratio	가사정보량 (문자당 2byte)
제시한 방안	1 (272.75 bytes)	2506 bytes
기존의 방안	110 (30072 bytes)	

### 참고문헌

- [1] Marco Furini and Lorenzo Alboresi, “Audio-Text Synchronization inside mp3 files: A new approach and its implementation” IEEE, 2004.
- [2] LG, DVD Karaoke Format Ver. 2.1.2.
- [3] Microsoft Corporation, Advanced System Format Specification Revision 01.20.03. 2004.