
소파변환을 사용한 오디오 데이터 베이스 검색 기반에서의 오디오 색인에 관한 연구

최귀열*, 곽철성**

A Study on Audio Indexing Using Wavelet Transform for Content-based Retrieval in Audio Database

Gwi-Yoel Choi, Chil-Seong Kwak

요 약

디지털 기술 발전에 따른 오디오 데이터의 증가는 여러 컴퓨터 응용에 사용되면서 데이터를 관리하고 사용하기 위해, 내용기반 질의와 유사성 검색과 같은 새로운 기능을 갖는 데이터베이스 시스템의 개발이 불가피하게 됐다. 내용기반 질의를 위한 빠르고 정확한 검색은 이러한 응용 시스템들에 필요하다. 효율적인 내용기반 색인과 유사성 검색의 설계는 관련성 있는 데이터의 빠른 검색을 제공하기 위한 주된 요소이다. 본 논문에서는 소파(Wavelet) 변환을 이용한 한국 전통 음악 데이터베이스의 오디오 색인을 위한 방법을 제안한다. 또한 소파 변환을 이용해 오디오 데이터에 대한 색인의 가능성을 보인다.

ABSTRACT

Amounts of audio data used in several computer application have necessitated the development of audio database systems with newer features such as content-based queries and similarity searches to manage and use such data. Fast and accurate retrievals for content-based queries are crucial for such systems to be useful. Efficient content-based indexing and similarity searching schemes are keys to providing fast and relevant data retrievals. This paper present a method for indexing of Korean Traditional Music audio data based on wavelets. Also this paper present possibility of wavelet based audio indexing.

* 재능대학 정보통신과 부교수

** 재능대학 전자과 부교수

접수일자 : 1999년 11월 18일

I. 서론

멀티미디어 데이터가 빠르게 증가하면서 데이터 베이스의 구축이 증가하였고 이러한 데이터들을 고속 검색하기 위하여 많은 검색 방법이 도입되었다. 그러나 오디오의 색인화에 대한 연구는 오디오의 데이터 특성상 다른 매체들의 색인화에 비해 주목할만한 연구가 없었다.

본 연구에서 우리는 일반적인 목적을 지원하는 데이터가 묘사에 접근하는 원래의 오디오로부터 색인의 유도에 초점을 맞춘다. 그리고 색인을 끌어내기 위해 소파 변환의 사용을 제안한다. 소파변환이 신호처리나 데이터 압축에서 연구, 사용되고 있더라도, 오디오 및 멀티미디어 데이터 베이스 색인화에는 넓게 적용되지 않았다.

전통적인 데이터베이스의 검색은 핵심어와 같은 문자열 기반의 색인을 사용한다. 오디오 데이터와 같은 멀티미디어 데이터를 위한 문자열 기반의 색인은 매우 제한적이다. 이것은 오디오 데이터의 일정하지 않은 성질과 데이터 표현의 다양성, 그리고 이것을 표현하기 위한 문자열 기반의 색인의 형식화가 어렵고, 데이터 표현이 주관적 판단에 근거하는 이유 때문이다. 따라서 내용기반 응용에서는 데이터의 효과적인 사용을 위해 다양한 질의 표현에 대응한 색인 구조가 요구된다. 본 논문에서는 단지 소파변환을 사용하는 색인 유도에 초점을 둔다.

II장에서는 기존의 내용기반 검색을 위한 오디오 색인에 대한 연구를 간략히 소개하며, III장에서는 색인을 위한 소파변환의 이용에 대하여 기술한다. IV장에서는 제안된 오디오의 내용기반 검색 방법을, V장, VI장에서는 실험에 의한 결과와 결론을 그리고 향후 연구방향에 대하여 서술한다.

II. 내용기반 검색을 위한 오디오 색인 연구

기존 데이터베이스의 전형적인 다 차원 색인구조의 자세한 내용은[5, 6, 7, 8]과 그것의 참조를 참고하기 바란다.

일반적으로 오디오를 위한 색인화 방법은 다음과 같이 분류 할 수 있다. 첫 번째로 신호의 통계에 기반한 색인화 방법으로 이 방법에서는 샘플들

의 평균, 편차, 영 교차(zero-crossing), 자기 상관, 히스토그램과 샘플들의 차이 같은 전체 데이터 또는 데이터 블록 중 어느 한쪽의 신호 통계가 사용된다. 다음으로 음향의 특성에 기반한 색인화로 이 방법은 오디오 신호의 측정할 수 있는 양을 사용하여, 음조 같은 음향 속성, 큰 소리, 명료성, 그리고 조화가 유도된다. 그 결과로 통계적인 분석(평균, 편차 등)은 색인화와 탐색에 사용되는 특징 벡터를 끌어내기 위해 적용된다. 그리고 변환에 기반한 색인화 방법으로 이 방법은 변환(DCT같은)이 데이터(블록)에 적용되어지고, 적당한 계수가 선택되어진다. 그리고 이것을 탐색과 검색을 위한 색인처럼 사용한다. 이것은 사용될 적절한 변환과, 블록의 크기, 유지되어야 할 계수의 수 등 고려되어야 할 요소가 있다. DCT나 소파 변환에 의한 색인방법은 이 같은 변환 기반의 방법에 속한다.

III. 소파 변환 기반의 오디오 색인

소파 변환은 주어진 데이터로부터 서로 다른 주파수 성분을 끌어내고 이것의 크기에 맞는 해상도를 갖는 성분을 구한다. 소파 분석에 사용되는 스케일이라는 용어는 주파수의 역, 또는 시간해상도와 일치한다.

크게 두 가지 기법이 있는데 (1)연속 소파 변환(CWT)과 (2)이산 소파 변환(DWT)이다. CWT에서는 선택된 분석 소파가 소파 함수에 의해 곱해지고 그 변환은 시간 축 신호의 분리된 다른 구획에서 따로따로 계산되어진다. 이것이 STFT 또는 DCT와 다른 소파 변환의 가장 큰 특징이다.

III-1. 소파 변환 계수 추출

DWT는 신호를 서로 다른 주파수 대역에서 신호분해에 의해 다른 해상도로 근사 계수와 상세 계수로 분석한다. DWT는 스케일링 함수와 소파 함수라 불리는 두 집합을 사용하며, 각각은 저역 통과 필터와 고역 통과 필터로 구성된다. 서로 다른 주파수 대역에서의 신호 분해는 시간 영역 신호의 연속하는 저역 통과와 고역 통과 필터링과 필터링 후의 다운 샘플링에 의한 신호를 얻는다.

그림 1은 DWT/ 3 계층 동안의 서브밴드 부호화 처리를 보이며, $H[n]$ 과 $G[n]$ 에서는 DWT에 의한 저역 통과와 고역 통과 필터를 사용한다. 본 논문의 소파 변환 기반 색인화 방법에서는 오디오 데이터의 분석을 위해 MATLAB의 WAVEDEC 함수를 사용한다. WAVEDEC는 신호의 다층 1-차원 소파 분해를 행한다.

출력 분해 구조는 Wavelet 분해 벡터 C와 부기 벡터 L을 포함한다. 분해 벡터 C는 DWT 계수의 집합으로 되어 있고, 근사 계수와 상세 계수로 구성 되어있다. 출력 분해 구조는 다음에 의해 분석 된다.

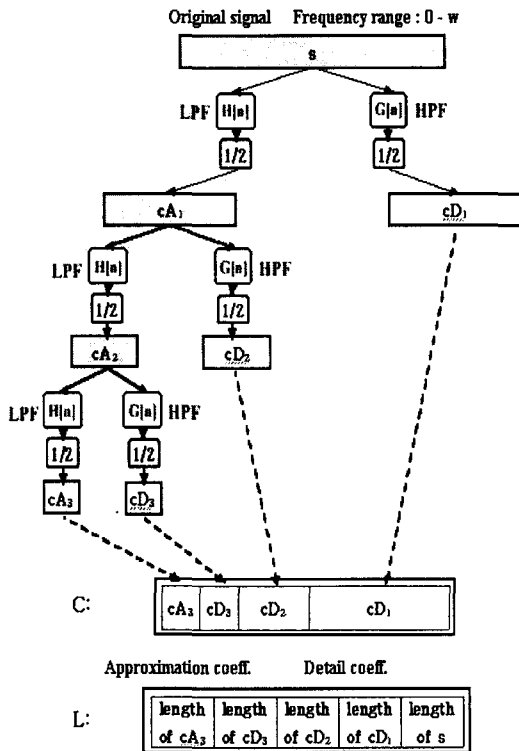


그림 1. DWT 서브밴드 분해 처리과정.
Fig. 1 DWT subband decomposition process.

$$C = [cA_N, cD_N, \dots, cD_1]$$

$$L = [length(cA_N), length(cD_N), \dots, length(cD_1), length(s)]$$

본 논문에서의 검색은 한국 전통 악기(이하 국악 악기)에 의해 독주된 오디오 데이터를 대상으로 한다. 국악 악기는 악기의 재료, 연주기법, 연주형태 등이 서양악기에 비해 다양하다. 또한 피아노와 같은 서양 악기는 정현파에 가까운 소리를 연주하나 가야금과 같은 국악 악기는 뜯거나, 밀거나, 튕기는 등, 연주 기법의 다양한 특성 때문에 악기의 정확한 분석알고리즘 구현이 서양악기에 비하여 용이하지 않다.

서양악기는 박자, 음정, 음압 등의 몇 가지 정보로도 연주가 가능하지만 국악 악기는 이외에도 다양한 연주 주법 표현이 요구되므로 연주에 필요한 정보가 훨씬 많다. 따라서 국악 오디오 데이터의 분석을 위해서는 보다 많은 정보를 포함할 필요가 있다.

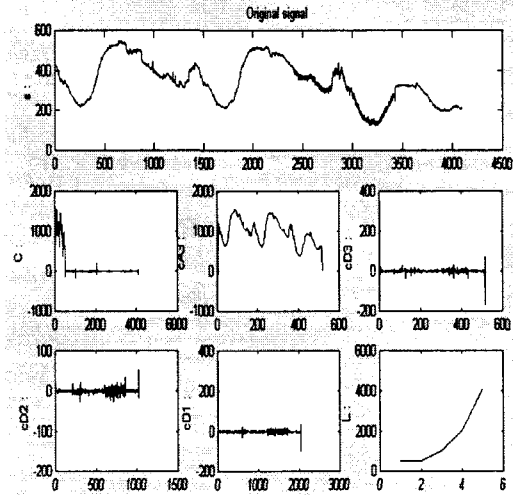


그림 2. 1-D DWT 계수.
Fig. 2 1-D DWT Coefficients.

본 논문에서는 병주나 합주 데이터의 분석은 고려하지 않았다. 합주 데이터의 경우는 더 많은 분석 알고리즘이 필요하며, 오디오 검색에서 두 악기 이상의 합주나 병주에 대한 연구는 아직 활발히 진전되지 못하고 있다.

소파 분해는 적당한 소파 필터를 사용하여 N번 데이터에 적용되었다. 그림 2는 임의의 데이터를 Daubechies 소파(db4)를 사용하여 계수 분리한 것

을 보이고 각각 s 는 원 신호 C 는 분해 벡터, L 은 부기 벡터, $cD1, cD2, cD3$ 는 각각 계층별 상세 계수이며, $cA3$ 는 근사 계수를 보인다.

그림 3과 그림 4는 국악 악기 중 가야금과 거문고에 의해 연주된 오디오 신호에 소파 분해를 적용한 예이다.

III-2. 색인을 위한 계수 선택

오디오를 소파 변환하면 낮은 해상도의 저주파 성분과 세 개의 고주파 성분을 얻는다. 저주파 성분은 소파계수의 대개의 에너지를 담고 있어 매우 중요하게 고려된다. 오디오를 소파 변환으로 분해함에 따라 상관성이 제거되지만 여전히 고주파 성분들 사이에 계층적 상관성이 존재한다. 따라서 색인으로 사용하기 위해 선택되어질 계수 고려에서 이러한 여분성을 제거하여 계수를 선택한다.

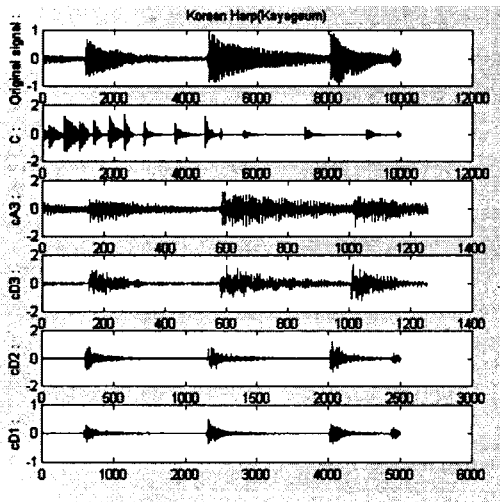


그림 3. 가야금 오디오 신호의 소파 분해.
Fig. 3 Wavelet decomposition of Kayageum audio.

그림 5는 그림 2와 같은 신호의 소파 변환에 의한 계수들의 상관성을 보인다.

소파 변환 계수의 색인 이용에는 여러 가지 방법을 생각해 볼 수 있다. 각각의 방법에서 근사 계수는 원 신호의 대개의 에너지를 갖는 저주파 성분들이므로 모두 사용한다. 그리고 각각의 방법에

서 상세 계수들의 색인을 위한 추출 방법이 고려된다.

첫 번째 방법으로는 각각의 상세 계수의 일정 개수를 색인을 위한 정보로 사용한다. 이 방법은 고주파 성분들의 에너지 분포를 고려하지 않았기 때문에 필요 없는 정보를 포함할 경우가 있다.

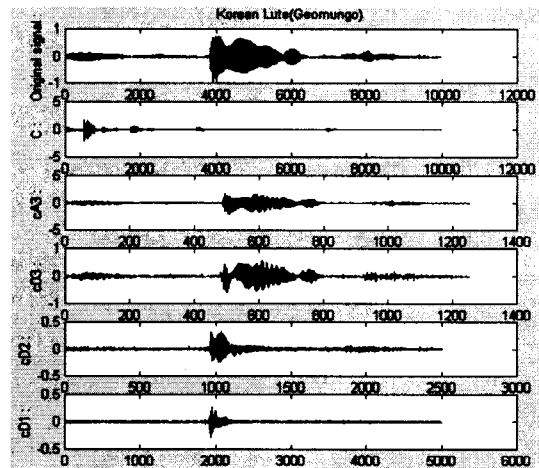


그림 4. 거문고 오디오 신호의 소파 분해.
Fig. 4 Wavelet decomposition Geomungo audio.

두 번째 방법은 각층의 상세 계수들로부터 에너지를 갖는 계수들의 일정량을 사용하는 것이다. 이 방법은 사용될 계수들의 위치정보를 함께 포함해야 하고, 분석되어야 할 데이터마다 위치 정보가 다르다.

$$C = [cA_N, cD_N, \dots, cD_1]$$

$$L(1) = \text{length}(cA_N),$$

$$L(j) = \text{length}(cD_{N-j+2}),$$

for $j=2, \dots, N+1$ and $L(N+2) = \text{length}(s)$

..... (2)

라 할 때,

첫 번째 방법에서 3 계층 분해($N=3$)이고 $n'a$ 와 $n'j$ 가 각각 선택된 근사 계수와 i 층의 상세 계수의 수를 의미한다고 할 때, 색인을 위하여 추출될 Cindex 와 Lindex 는 다음과 같다.

$$C_{index} = [C(1:n'_a), C(L(1)+1:L(1)+n'_3), C(L(1)+L(2)+1:L(1)+L(2)+n'_2), C(L(1)+L(2)+L(3)+1:L(1)+L(2)+L(3)+n'_1)]$$

$$L_{index} = [(L(1), n'_a), (L(2), n'_3), (L(3), n'_2), (L(4), n'_1), L(5) \dots \dots \dots (3)$$

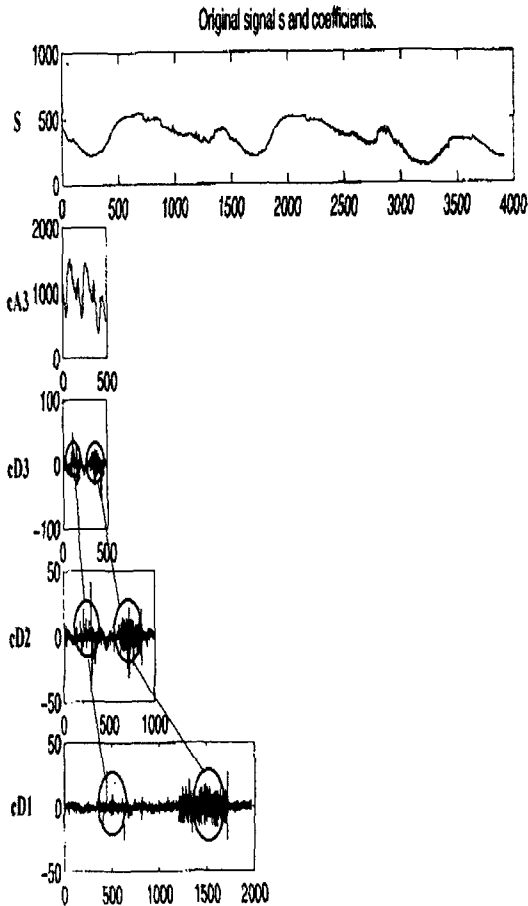


그림 5. DWT 계수의 특성.
Fig. 5 Characteristic of DWT coefficients.

두 번째 방법에서는 C'의 상세 계수들은 원래 C의 서로 다른 층에서의 최고 값들의 일정한 개수(색인의 크기를 맞추기 위하여)를 선택한다. 그러므로 각 층에서 선택된 계수들의 위치는 일정하지 않다. 따라서 선택된 계수들의 위치정보를 Lindex에 포함시킨다. p1, ... pni를 L에서 i층에서의 상세 계수들의 위치를 의미한다고 하면, Lindex는 다음

과 같다.

$$L_{index} = [(L(1), n'_a), (L(2), n'_3, p_1, \dots, p_{n_3}), (L(3), n'_2, p_1, \dots, p_{n_2}), (L(4), n'_1, p_1, \dots, p_{n_1})] \dots \dots \dots (4)$$

본 논문에서 제안한 방법은 두 번째 방법에서 상세 계수들의 상관성을 이용하여 저주파성분에 가장 가까운 상세 계수의 전체를 사용함으로써 다른 층 상세 계수들의 최고 값들과 최고 값들의 일정량을 추출하기 위한 계산을 해야하는 것과 위치정보를 포함해야하는 것을 없앴다. 또한 서로 상관성 있는 위치의 계수를 반복하여 검사하는 계산을 없앴다. 이를 위해서 제안한 방법의 계수 선택은 단지 cA3와 cD3를 갖는 Cindex와 그리고 부가 정보(위치정보, 샘플크기)를 위하여 Lindex를 사용한다.

예를 들어 3 계층 분해(N=3)를 고려해 볼 때, 제안한 방법에서 색인을 위하여 사용될 Cindex와 Lindex는 다음과 같다.

$$C_{index} = [C(1:L(1)), C(L(1)+1:L(2))] \dots \dots (5)$$

$$L_{index} = [L(1), L(2)]$$

그림 6은 앞에서 설명한 색인을 위한 계수 추출 개념을 보인다.

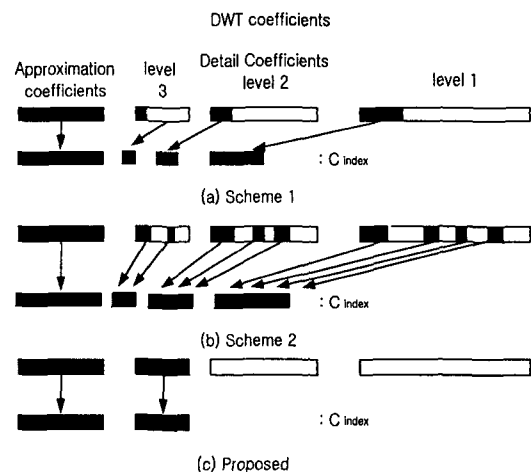


그림 6. DWT 계수로 부터의 색인 추출.
Fig. 6 Index extraction from DWT coefficients.

계수는 탐색과 검색에 사용되기 위해 선택된다. 색인 생성을 위한 의사코드와 추출은 그림 5에 주어진다.

그림 7의 알고리즘은 *.wav 파일을 입력으로 받아 헤더정보를 제거하고 실제 데이터를 소파 변환하여 색인을 위한 계수정보를 저장하는 과정을 나타내고 C 언어에 의해 헤더정보 제거 및 실제 데이터의 추출을 구현하고, 저장된 실제 데이터는 MATLAB 의 1차원 소파 변환에 의해 계수 값을 얻는다.

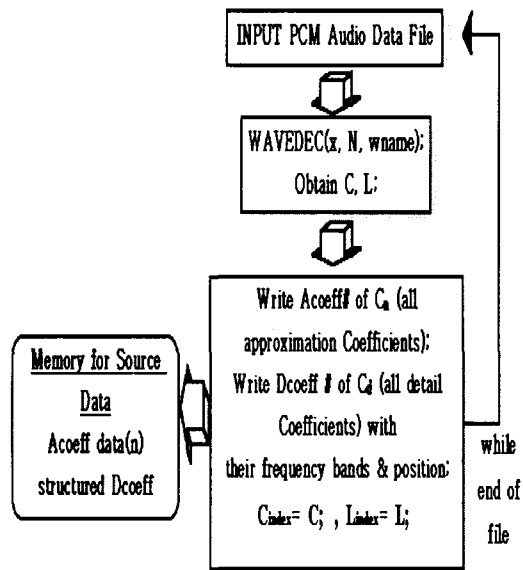


그림 7. 색인을 위한 계수선택 알고리즘 흐름도
Fig. 7 Coefficient selection for indexing algorithm flowchart.

색인을 위한 계수의 추출은 계수들의 상관성이 고려되며, 이것은 색인을 위한 계수 정보의 중복을 피하고 색인을 위한 데이터의 크기를 줄이기 위함이다.

IV. 제안된 오디오의 내용기반 검색

그림 8의 알고리즘은 이미 저장된 색인들과 입력된 현재의 데이터와의 유사성을 검색하여 검색

결과를 출력 각각의 리스트를 출력할 수 있도록 한다. 또한 소파 변환에 의해 획득된 색인들에 기반한 탐색을 설명한다.

이것은 일반적인 탐색이고 구조적인 색인들의 데이터구조에 특정한 동작을 고려하지 않는다는 점을 주의해야한다. 따라서, 여기에서의 주목적은 오디오 데이터를 위한 색인으로 소파 변환 계수의 사용 가능성을 입증하기 위함이다.

제안된 소파 변환 방법에 의한 색인 기법은 Intel Pentium-II 400 PC에서 MATLAB과 C 언어를 이용하여 구현하고, 9종류 악기(거문고, 단소, 양금, 대금, 해금, 당적, 가야금, 피리, 사물놀이 악기)의 100개 이상의 국악 오디오 데이터를 대상으로 실험하였다.

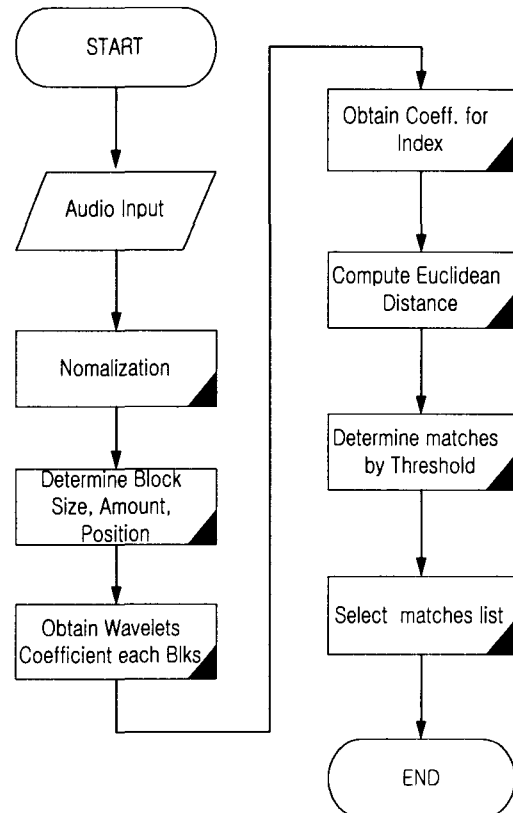


그림 8. 검색 흐름도.
Fig. 8 Searching flowchart

V. 실험

각 데이터는 .wav 형식으로 44바이트의 헤더정보를 가진다. 그리고, 샘플링은 8Khz, 8-bit, Mono를 사용하였다.

DWT 계수를 끌어내기 위해 Mother Wavelet으로 Daubechies-4를 사용한다. 블록의 크기는 고정 크기(512 sample.)로 하고 블록 수는 10과 20일때, 실험은 10개의 예에 의한 질의를 대상으로 이루어졌으며, 질의에 약간의 잡음이 더해졌다. 소파 계수 추출을 위한 데이터들은 각각 정규화 과정을 거친다.

모든 방법에서 적합성은 유클리드의 거리가 사용되었으며 측정식은 다음과 같다.

$$D_{xy} = \sum_{i=1}^n (x_i - y_i)^2 / [(\sum_{i=1}^n x_i^2) \cdot (\sum_{i=1}^n y_i^2)]^{1/2} \dots (6)$$

미터법의 성능은 검색정밀도 이고 우리는 이것을 검색 결과 출력의 검색 정확성의 확률(정밀도)로 비교하였고 정의는 다음과 같다.

$$\text{정밀도} = \frac{\text{적합한 결과물의 수}}{\text{총 검색 결과물의 수}} \times 100(\%) \dots (7)$$

표 1. 검색 정밀도의 비교
Table 1. Comparison of retrieval precision.

Qry.	Precision	
	10 blocks Pre.(%)	20 blocks Pre.(%)
1	13.04	14.30
2	28.57	66.66
3	49.13	60.00
4	11.15	33.33
5	30.00	66.66
6	66.57	59.11
7	42.86	40.00
8	40.00	66.66
9	33.33	42.86
10	40.00	66.66

실험 결과 기존의 색인을 위한 계수 추출 방법과 거의 같은 품질의 검색 정밀도를 보였다. 또한

분해 벡터 전체를 색인으로 사용하여 검색했을 경우와 거의 비슷한 결과가 나오는 것을 실험적으로 알 수 있었다.

표 1의 검색 정밀도 비교는 소파 기반의 색인 방법에서 하나의 오디오 데이터 파일, 즉 8 kHz × 10 sec 의 80,000개의 고정 크기의 샘플을 갖는 파일을 각각 512개 샘플 개수를 갖는 블록의 개수가 파일당 10개와 20개일때의 정밀도이고, 대체로 만족할 만한 수준은 아니나, 블록의 개수가 늘수록 높은 정밀도를 갖는다. 본 논문에서는 일정한 크기의 오디오 데이터를 사용하였다 그러나 실제의 데이터베이스에서는 각각의 오디오 데이터의 크기가 다르므로 차후 실험에서는 가변적인 파일 크기에 대응하는 색인 구조를 가져야 하겠다.

VI. 결론

멀티미디어 데이터베이스에서 오디오 데이터의 내용에 근거한 정확한 검색을 수행하기 위해서는 적절한 색인들의 유도가 필요하다. 각각의 색인들은 오디오 데이터의 특징을 고려하여 구조적으로 나타낼 필요가 있다.

본 논문에서는 예제 기반의 질의를 위한 오디오 데이터베이스에서의 오디오 데이터 색인을 위해 소파를 이용한 방법을 제안했다. 또한, 실험적으로 소파 변환이 국악 오디오 데이터베이스에서의 내용 기반 검색에 사용될 수 있음을 보였다.

향후 내용기반 오디오 검색을 위한 연구 방향으로서는 더 정확한 검색을 제공하는 유연한 색인들의 다른 소파 필터들에 대한 연구와 국악 오디오 데이터와 같은 변화가 심한 데이터의 조작과 짧은 질의를 위한 연구가 필요하며, 나아가 소파와 수반하는 다른 변환 방법과의 병행에 관한 연구가 필요하다.

참고 문헌

[1] E.Wold et al. Content-based classification, search and retrieval of audio data. IEEE Multimedia Magazine, 1996.
[2] A.Ghias et al. Query by humming. Proc. ACM

- Multimedia Conf., 1995.
- [3] S.R.Subramanya et al. 'Transform-based Indexing of Audio Data for Multimedia Databases', IEEE Int'l Conference on Multimedia Systems, Ottawa, June 1997.
 - [4] S.R.Subramanya 'Experiments in Indexing Audio Data' Tech. Report, GWU-IIST, January 1998.
 - [5] N.Beckmann et al. 'The R*-tree: An Efficient and Robust Access Method for Points and Rectangles' pp322-331, ACM SIGMOD, May 1990.
 - [6] D.B.Loment and B.Salzberg 'The hb-tree: A Multiattribute indexing method with good guaranteed performance' ACM TODS, 15(4), pp625-658, Dec. 1990.
 - [7] J. T. Robinson 'the k-D-B-tree: A Search Structure for Large Multidimensional Dynamic Indexes', Proc. ACM SIGMOD, pp10-18, 1981.
 - [8] J.Nievergelt et al. 'The The Grid File: An Adaptable Symmetric Multi-key File Structure', ACM Trans. on Database Systems, Vol. 9, No. 1, 1984.
 - [9] M. Vetterli et al., Wavelet and Subband Coding, Prentice-Hall PTR, New Jersey 1995.



최 귀 열(Gwi-Yoel Choi)
1980년 2월 숭실대학교 전자공
학과 졸업(공학사)
1987년 8월 숭실대학교 대학원
전자공학과 졸업(공
학석사)

2000년 2월 동대학원 박사과정 수료
1991년 3월 - 현재 재능대학 정보통신과 부교수
2000년 10월 -현재 해양정보통신학회 종신회원
*관심분야 : DSP 시스템, VOD 시스템, 신경망 및
퍼지



곽 칠 성(Chil-Seong Kwak)
1981년 광운대학교 응용전자공
학과 졸업(공학사)
1986년 동대학원 전자통신공학
과 졸업(공학석사)
1991년 3월 - 현재 재능대학

전자과 부교수
*관심분야 : 응용전자 회로 설계, 마이크로 프로세
서 응용설계