

마이크 성능과 노래 만족도가 허밍질의에 미치는 영향†

현경우*, 남현우*, 박능수*, 권순일**
*건국대학교 컴퓨터공학부
**한국과학기술연구원
e-mail : xacdo@xacdo.net

How Microphone Performance and Sung Confidence Influence on Query by Humming

Kyungwoo Hyun*, Hyunwoo Nam*, Neungsoo Park*, Soonil Kwon**
*Dept. of Computer Science, Konkuk University
**Korea Institute of Science and Technology

요 약

이 논문은 허밍질의의 정확도 향상에 마이크와 같은 장비의 특성, 허밍을 하는 사용자의 숙련도가 어느 정도 반영되는지 알아본다. 연구의 결과 검색의 정확도는 마이크 성능에 독립적임을 보였고, 사용자가 노래를 아는 정도와 노래를 부른 만족도가 검색 순위를 상승시키는데 미치는 영향이 없거나 미미함을 실험 결과를 통해 보았다. 이와 같은 실험 결과들은 허밍질의의 검색의 정확도 향상을 위한 연구에서 참고자료로 사용될 것이다.

1. 서론

음원 서비스 업체 멜론(MelOn)이 약 120 만 곡 (2007 년 5 월)[1], 미국 iTunes Store 가 약 8 백만 곡 (2008 년 9 월)을 보유하고 있으며[2], 앞으로도 디지털 음원은 계속 증가할 것으로 보인다. 이렇게 많은 노래를 사람의 힘으로 제공하는 데는 한계가 있으므로, 이를 자동화 시켜줄 시스템이 필요하다. 그 방법 중 하나로 허밍질의(Query by Humming, QbH)가 있다.

허밍질의는 “흠흠흠~ 나나나~” 하는 사람의 허밍으로 음악을 검색하는 방법으로, 음악정보검색(Music Information Retrieval, MIR)의 한 연구분야이며 MPEG-7 표준에도 포함되어 있다.[3] 제목, 음악가, 가사 등의 명시적인 데이터로 찾지 않고(not text-based search), 실제 내용으로 찾는다(context-based search).

허밍질의의 검색 정확도는 높다. 2007 년 MIREX(Music Information Retrieval Evaluation eXchange)의 Query-by-Singing/Humming 부문에서 1 위를 한 시아오 우, 밉 리(Xiao Wu, Ming Li)는 통제된 환경에서 92.5%, 93.7%의 정확도를 보였으며[4], 2008 년 Erdem Unal 의 실험에서는 25 년 이상 숙련된 음악가의 경우 88% 이상의 정확도를 보였다.[5]

현재 미국의 미도미(Midomi)[6]가 허밍질의를 서비스하고 있고, 한국의 벅스(Bugs)는 나요미디어를 인수하여 2007 년 10 월 허밍서치(humming search)를 제공했다.(현재는 중단)[7]

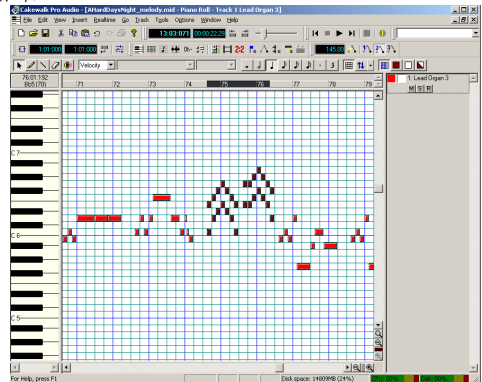
이와 같이 허밍질의는 높은 정확도를 바탕으로 대중화 단계에 있다. 이 논문은 앞으로 휴대폰 등의 일상적인 생활 기기에서 낮은 품질로 녹음해도 허밍질

의가 가능한지 기존의 연구를 바탕으로 실험을 통해 검증했다.

2. 실험 설계

실험의 보편성을 위해, 대중음악으로 유명한 비틀즈(The Beatles)의 노래를 선택했다. 비틀즈의 미디(MIDI) 파일 206 곡을 ABC 순서대로 배열하여, 앞에서부터 20 곡을 추려서 실험에 사용했다.

이 20 곡에서 멜로디에 해당하는 트랙(track)를 잘라내고, 그 멜로디 중에서 인장적이라고 생각되는 4 마디에서 16 마디 길이의 주제(theme)를 선택하여 잘라냈다. 그리고 화음으로 구성된 미디 파일들은 단음으로 바꾸었고, 수작업으로 주제를 잘라내는 작업들은 하였다.



(그림 1) 화음을 단음으로 바꾸는 작업

† 본 연구는 서울시 산학연 협력사업(10581)의 일환으로 수행하였음.

이렇게 만들어진 20 개의 단음 음악 데이터와 20 개의 주제로 나누어진 음악 데이터를 생성했다. 추가적으로 각각 비슷한 분포의 잡음(noise dummy)이 포함된 200 곡의 더미 음악 데이터들을 랜덤으로 생성하여, 총 220 곡 중 20 곡을 찾아내는지를 실험했다.

컴퓨터는 2 대를 사용했다. 하나는 노트북 컴퓨터의 내장 마이크를 사용했고, 다른 하나는 데스크탑 컴퓨터에 USB 오디오 인터페이스(Audio Interface)와 고급 마이크(Shure SM58)를 사용했다.

총 3 명의 피험자에게 원곡을 들려준 후(2~4 분), 노래를 아는지 여부를 기록하고, 주제를 들려주고(5~20 초), 주제를 다시 이어폰으로 들으면서 노래를 따라 부르게 했다. 이 노래를 2 개의 컴퓨터에서 동시에 녹음하여 기록했다. 그리고 피험자에게 자신이 부른 노래의 만족도를 평가하게 했다. 이와 같은 방법으로 총 20 곡을 녹음했다.

<표 1> 인지도 척도

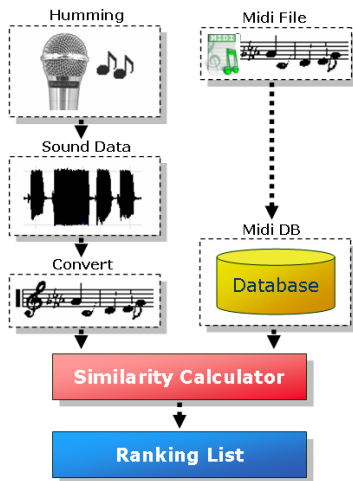
1	전혀 모른다.
2	약간 모른다.
3	알쏭달쏭하다.
4	약간 안다.
5	잘 안다.

<표 2> 만족도 척도

1	아주 못 불렀다.
2	약간 못 불렀다.
3	그럭저럭 불렀다.
4	약간 잘 불렀다.
5	아주 잘 불렀다.

3. 허밍 검색 시스템 구조

이번 논문에서는 UDR 멜로디 표현법을 개선한 그립허밍[10] 허밍 검색 시스템을 사용한다. 이 시스템은 크게 허밍에서 신호처리를 통해 음(note)을 추출하는 과정, 미디 파일에서 음을 추출하는 과정으로 나눌 수 있다.



(그림 2) 허밍 검색 시스템 구조

그렇게 추출된 각 음원의 멜로디 데이터들은 Similarity Calculator 에서 Edit Distance 알고리즘을 사용하여 유사도를 계산한다. 검색의 최종 결과로는 유사도 순으로 순위를 매겨 출력한다.

추가적으로 이번 논문에서 수행할 여러 실험들과 결과들은 그림 2 의 시스템에서의 경우를 전제로 하고 있다. 따라서 본 시스템 외의 알고리즘이 적용된 시스템에서는 동일한 결과를 보여주지 않을 수 있다.

4. 오비오(Aubio)의 음 결정 방법

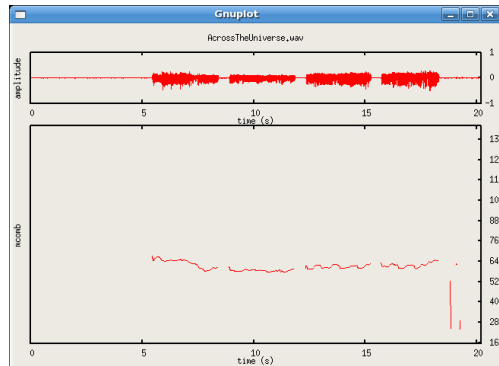
3 장에서 살펴본 허밍 검색 시스템에서 녹음한 파일들은 오비오(Aubio) 라이브러리에서 제공하는 도구인 aubionotes 를 사용하여 음(note)의 음정(pitch)과 지속시간(duration)을 추출했다.[8]

오비오는 4 단계를 거쳐 음을 추출한다. 첫 번째는 잡음 제거(noise gate)로, 역치(threshold, 기본값은 10%) 미만의 작은 음은 잡음으로 간주하고 무시한다. 두 번째는 시작음 탐지(onset detection)으로, 음이 시작하는 부분을 탐지하여 분할(segmentation)하며, 기본값으로 쿨백 라이블러 분기(Kullback Liebler divergence)와 복소수 영역(complex domain) 탐지를 사용한다.



(그림 3) 시작음 탐지(onset detection)

세 번째로 음정 추측(pitch estimation)은 사람이 부른 음이 기반음정이 없어질 수 있기 때문에(missing fundamental) 나머지 음정들을 가지고 추측한다(auto-correlation).

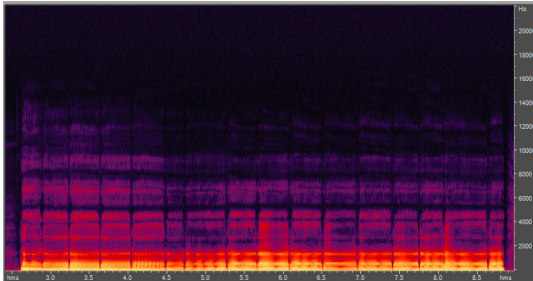


(그림 4) 음정 추측(pitch estimation)

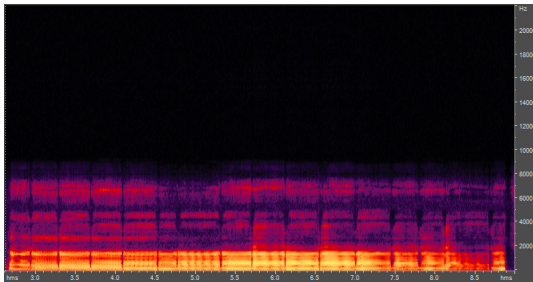
마지막으로 위에서 탐지하고 추측한 시작 음과 음정을 가지고 음을 결정한다(note decision).[9]

이렇게 네 단계를 거치면서 잡음(noise), 왜곡(distortion)에 강하도록 각각의 단계마다 역치(threshold)값을 두어 자잘한 값을 잘라내고, 데이터를 메타데이터로 단순화하므로, 마이크 성능에 따른 차이는 없는 것으로 보인다.

5. 마이크 성능에 따른 검색 성능 비교

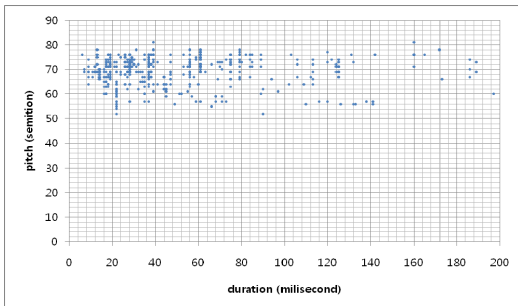


(그림 5) 피험자 3 이 고급 마이크로 부른 Across The Universe 의 스펙트로그램(spectrogram)

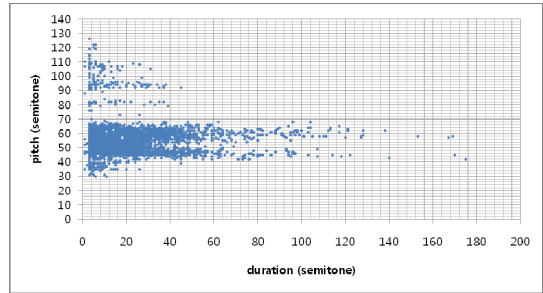


(그림 6) 피험자 3 이 저급 마이크로 부른 Across The Universe 의 스펙트로그램(spectrogram)

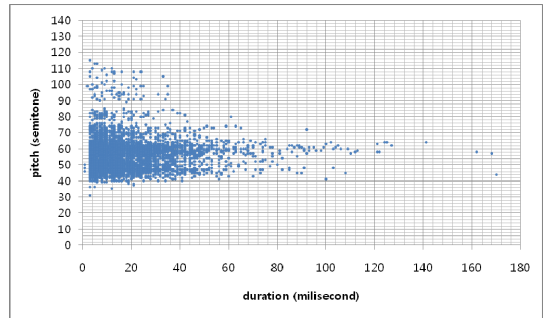
그림 5 와 그림 6 을 살펴보면 고급 마이크의 경우 좀 더 넓은 음역을 선명하게 표현하고 있으나 물결모양의 음성 지문, 에너지의 높낮이 등에는 큰 차이가 없다.



(그림 7) 원본 미디 DB 의 음 분포



(그림 8) 고급 마이크의 음 분포



(그림 9) 저급 마이크의 음 분포

그림 7, 그림 8, 그림 9 을 보면 고급 마이크가 저급 마이크보다 조금 조밀한 분포를 보여주지만 전체적으로는 큰 차이가 없다.

이 데이터들을 유사도를 계산하는 Edit Distance 알고리즘으로 비교하였을 때, 검색된 평균 순위는 표 3 와 같다.

<표 3> 검색 결과 평균 순위

실험자	종류	
	고급마이크	저급마이크
피험자 1	9.578947	9.947368
피험자 2	10.89474	13.21053
피험자 3	11	11.78947
평균	10.49123	11.64912

피험자가 부른 곡이 검색 결과에 오른 순위가 고급 마이크는 평균 10.49 위, 저급마이크는 11.65 위로, 전체 220 곡 중에 1.16 위, 0.53%의 미미한 차이가 있었다.

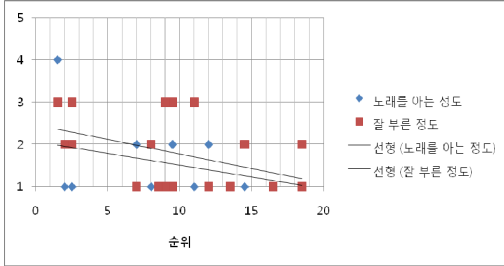
6. 피험자의 사전 지식, 만족도의 영향

이번 실험은 사용자가 부른 허밍의 만족도나 잘 아는 정도의 차이가 허밍검색의 정확도에 어떤 영향을 미치는지 알아본다.

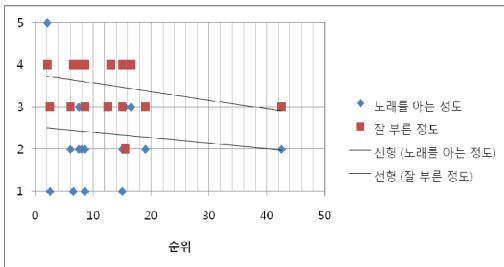
여기서 만족도나 인지도라는 척도는 주관적인 판단이 개입할 여지가 있어서 기존 논문에서는 객관적인 지표로 직업 성악가의 경력을 사용했다.[5] 하지만 일반 전체 인구 중 성악가가 차지하는 비중이 작으므로 일반 대중에게 이 척도를 적용하기는 좋지 않다.

이번 실험에서는 일반 대중에게 적용 가능한 척도

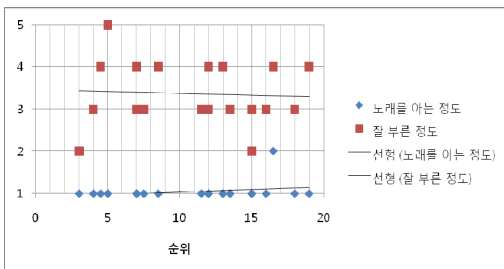
로서 피험자가 스스로 자신의 사전 지식과 만족도를 1에서 5까지의 척도로 자기 평가(self-rating)하도록 했다. 다음의 그림들은 허밍을 할 때 작성된 설문조사 결과를 통계적 수치로서 보여준다.



(그림 10) 피험자 1



(그림 11) 피험자 2



(그림 12) 피험자 3

이 결과는 피험자의 사전지식과 만족도가 검색순위에 미치는 영향이 없거나 미미함을 보여준다. 그림 10, 11을 보면 노래를 잘 알면 알수록 검색의 순위가 약간 높아지지만 그 정도는 미미했고, 자신이 잘 불렀다고 평가하는 정도도 마찬가지였다. 또한 그림 12을 보면 피험자 3의 경우 대부분의 노래를 모른다고 답했기 때문에 상관관계가 거의 없는 것을 확인하였다.

7. 결론

마이크 성능에 따른 차이가 없는 것은 아니지만 미미했다. 좋은 마이크가 음성을 정확하게 잡아주고, 노이즈가 적고, 왜곡이 적은 탓도 있지만, 허밍질에서 사용하는 알고리즘이 이런 부정확함, 노이즈, 왜곡에 강하도록 설계되었기 때문에 차이가 적게 난다고 볼

수 있다.

또한 주관적인 척도로 피험자의 사전 지식과 만족도가 검색 순위에 미치는 영향이 없거나 미미함을 확인하였다.

그러므로 향후 허밍질의 연구는 기기의 성능이나 개개인의 특성의 차이보다는 왜곡에 강한 알고리즘을 찾는 데 집중해야 할 것이다.

8. 향후 연구 과제

현재 허밍질의 알고리즘이 많은데 이를 벤치마킹할 표준 환경이 MIREX 외에는 없다. 음원 데이터베이스를 구축하는 작업은 노동집약적(labor-intensive)이고, 노래의 저작권 문제로 많은 자분이 여러 연구에 중복 투자될 여지가 있다. 그러므로 한국 환경에 적합한 허밍질의 표준 데이터베이스와 질의 예제 파일 등의 테스트 환경을 구축하여 알고리즘 간의 우위를 쉽게 비교할 수 있도록 하여 허밍질의 연구의 객관적 지표를 세우고 동기부여를 할 필요가 있다.

그리고 허밍질의 연구가 정확한 검색결과에만 치중하고 있는데, 정확도와 함께 심리적인 만족도, 자신감 등의 사용자 경험을 측정할 객관적인 척도를 개발하는 것도 필요하다. 또한 허밍질을 했을 때 자신이 생각했던 곡과 검색결과와의 유사점을 쉽게 파악할 수 있는 유저 인터페이스 환경의 개발도 필요하다.

참고문헌

- [1] 전자신문 2007년 5월 7일 기사 “SKT, 영업장 전용 비즈메일 서비스 출시”
- [2] <http://en.wikipedia.org/wiki/Itunes>
- [3] http://en.wikipedia.org/wiki/Query_by_humming
- [4] http://www.music-ir.org/mirex/2007/index.php/Query-by-Singing/Humming_Results
- [5] Erdem Unal et al, “Challenging Uncertainty in Query by Humming Systems: A Fingerprinting Approach”, IEEE Transaction on Audio, Speech, and Language Processing Vol. 16, No. 2, February 2008
- [6] <http://www.midomi.com/>
- [7] http://notice.bugs.co.kr/notice/bugsnotice_view.asp?Gubun=Notice&SeqNo=1078
- [8] <http://www.aubio.org/>
- [9] Paul M. Brossier, “Fast labeling of notes in music signals”, in Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004), Barcelona, Spain, October 10-14, 2004.
- [10] 남현우, “허밍 질의를 위한 그룹 허밍 표현법”, 한국정보처리학회 추계학술발표대회 제 14 권 2 호, 2007.