비디오 의미 파악을 위한 멀티미디어 요약의 비동시적 오디오와 이미지 정보간의 상호 작용 효과 연구

A Study on the Interactive Effect of Spoken Words and Imagery not Synchronized in Multimedia Surrogates for Video Gisting

김 현 희(Hyun-Hee Kim)*

목 차

초 록

본 연구는 오디오 및 이미지 정보가 비동시적으로 결합된 오디오/이미지 요약이 오디오 요약 또는 이미지 요약만 사용했을 때 보다 어떤 상호 작용 효과를 가지고 있는지 살펴보았다. 이를 위해서 오디오/이미지 요약, 오디오 요약 및 이미지 요약을 비디오의 의미 추출에 있어서의 정확도 즉, 요약문 및 항목 선택의 정확도와 이용자들의 이 세 가지 요약에 대한 관점을 비교, 분석하였다. 분석 결과, 요약문 정확도에서는 비디오 유형에 관계없이 상호 작용 효과를 확인하였으나 항목 선택의 정확도에서는 상호 작용 효과가 입증되지 못했다. 끝으로 이용자들은 오디오/이미지 요약에 대해 오디오와 이미지 정보를 병행하여 시청함으로서 비디오 내용에 대한 이해를 빠르게 하지만 때로는 이 두 정보간의 비동시성으로 인하여 비디오 의미 파악을 방해하는 경우도 생겨난다고 기술하였다.

ABSTRACT

The study examines the interactive effect of spoken words and imagery not synchronized in audio/image surrogates for video gisting. To do that, we conducted an experiment with 64 participants, under the assumption that participants would better understand the content of videos when viewing audio/image surrogates rather than audio or image surrogates. The results of the experiment showed that overall audio/image surrogates were better than audio or image surrogates for video gisting, although the unsynchronized multimedia surrogates made it difficult for some participants to pay attention to both audio and image when the content they present is very different.

키워드: 멀티미디어 요약, 오디오/이미지 요약, 오디오 요약, 스토리보드, 비디오 검색, 이미지 요약
Interactive Effect, Unsynchronized Multimedia Surrogate, Video Gisting

* 명지대학교 문헌정보학과 교수(kimhh@mju.ac.kr)
논문접수일자: 2011년 4월 18일 최초심사일자: 2011년 4월 18일 게재확정일자: 2011년 5월 11일
한국문헌정보학회지, 45(2): 97-118, 2011. [DOI:10.4275/KSLIS.2011.45.2.097]

1. 서 론

1.1 연구 배경과 목적

비디오는 대용량 자원이기 때문에 비디오 전체를 보기 전에 적합성 판정을 위한 좀 더 세밀한 브라우징 과정이 필요하다(Kristin et al. 2006). 또한 영상물 자료에 대한 적합성 판정 유형은 이용자에 따라 많은 차이를 보이고 있고 때문에 비디오 자료를 검색하고 걸러내기 위해서 다양한 메타데이터와 멀티미디어로 구성된 대용물(이하, 멀티미디어 대용물)이 필요하다고 주장하기도 한다(Yang 2005).

전통적으로 비디오 검색은 텍스트 기반 메타데이터를 이용해 오고 있었으며 최근에 와서는 비디오의 주요 프레임들을 보여주는 이미지(비주얼) 요약인 스토리보드도 이용되고 있다. 그러나 텍스트 기반 메타데이터나 이미지 요약만으로 영상물의 의미를 파악하는데 한계가 있을 뿐만 아니라 비디오 특성 또는 장르에 따라서 메타데이터나 대용물의 효과가 달라질 수 있다고 생각한다.

일반적으로 가장 이상적인 방법 중 하나는 텍스트 기반 메타데이터와 함께 멀티미디어 대용물을 사용하는 것이다. Song과 Marchionini (2007)는 오디오 요약과 이미지 요약을 결합한 오디오/이미지 요약의 효과가 오디오 또는 이미지 요약만을 단독으로 사용했을 때 보다 더 효율적이라고 주장하고 있다. 또한 이들은 이용자들은 오디오 요약을 통해서 비디오 내용을 확인하고 추가적인 가치를 얻는다고 주장하고 있다. 이외에 오디오 정보가 중복된 정보를 제공함으로써 영상 정보를 보완한다는 주장도 있고 (Gunther et al. 2004), 오디오 정보가 영상 정보 보다 기억하고 정보를 이해하는데 더 효율적이라고 주장하기도 한다(Schmandt & Mullins 1995). 오디오 및 비디오 정보가 동시적으로 결합된 비디오 스킴은 전체 비디오 내용을 효과적으로 표현할 수 있어서 이용자의 이해도면에서 더 유리한 요약이 될 수 있다. 그러나 비디오 스킴을 실제 정보 검색 환경에서 사용하기에는 비용이 많이 들고 아직까지 기술적인 어려움이 있다.

Marchionini et al.(2009)은 오디오 및 이미지 정보가 비동시적으로 결합된 오디오/이미지 요약을 제안하였다. 그러나 이들은 이용자들이 이미지와 오디오 내용이 일치하지 않아서 비디오 의미 파악을 하는데 전혀 도움을 주지 않는 경우도 있다고 보고하고 있다.

따라서 오디오 및 이미지 정보가 비동시적으로 결합되어 구성된 오디오/이미지 요약이 비디오 검색 및 브라우징 과정에서 어떤 효과를 내고 있는지에 대한 체계적인 연구가 필요한 시점이다. 본 연구의 목적은 오디오/이미지 요약이 오디오 요약 또는 이미지 요약만 사용했을 때 보다 어떤 상호 작용 효과를 가지고 있는지, 또한 비디오 유형에 따라서 이러한 상호 작용 효과가 어떻게 달라지는지 살펴봄으로써 디지털 도서관 또는 모바일 환경에서 비디오 자료의 검색 및 브라우징의 효율성을 높이기 위한 방안을 모색하는데 있다.

1.2 연구 문제와 가설

본 연구는 오디오 및 이미지 요약의 상호 작용 효과를 분석하고자 오디오/이미지 요약, 오

디오 요약 및 이미지 요약을 비디오의 의미 추출에 있어서의 정확도, 즉 요약문 및 항목 선택의 정확도와 이용자들의 이 세 가지 요약에 대한 관점을 서로 비교하고자 한다. 요약문 정확도는 이용자들에게 멀티미디어 요약을 시청하게 한 후 각 비디오에 대한 요약문을 기술하게 한 후 평가한다. 이외에 항목 선택 정확도는 10개 항목들을 제시한 후 각 항목이 비디오 내용을 맞게 기술한 것인지 아니면 틀리게 기술한 것인지를 체크하도록 한다. 구체적으로 연구 문제와 가설을 기술하면 다음과 같다.

첫 번째 연구 문제는 전체 표본 비디오 집단(강의 및 연설 등 스크립트 중심의 비디오 유형 1과 영상 중심의 비디오 유형 2)에서 오디오/이미지 요약이 오디오 요약 또는 이미지 요약과 비교하여 비디오 의미 추출에 있어서의 정확도가 어떤 차이를 보이는가이다. 이러한 연구 문제는 다음과 같은 연구 가설 1과 2를 이끌어낸다.

- 연구 가설 1: 전체 비디오(비디오 유형 1과 2)에서 오디오/이미지 요약, 오디오 요약 및 이미지 요약에 의한 요약문 정확도에 차이가 있을 것이다.
- 연구 가설 2: 전체 비디오(비디오 유형 1과 2)에서 오디오/이미지 요약, 오디오 요약 및 이미지 요약에 의한 항목 선택의 정확도에 차이가 있을 것이다.

두 번째 연구문제는 비디오 유형(비디오 유형 1과 비디오 유형 2)에 따라서 오디오/이미지 요약이 오디오 요약 또는 이미지 요약과 비교하여 비디오 의미 추출에 있어서의 정확도가 어떤 차이를 보이는가이다. 이러한 연구 문제는 다음과 같은 네 개의 연구 가설 3-6을 이끌어낸다.

- 연구 가설 3: 비디오 유형 1에서 세 가지 요약에 의한 요약문 정확도에 차이 가 있을 것이다.
- 연구 가설 4: 비디오 유형 1에서 세 가지 요약에 의한 항목 선택의 정확도에 차이 가 있을 것이다.
- 연구 가설 5: 비디오 유형 2에서 세 가지 요약에 의한 요약문 정확도에 차이 가 있을 것이다.
- 연구 가설 6: 비디오 유형 2에서 세 가지 요약에 의한 항목 선택의 정확도에 차이 가 있을 것이다.

세 번째 연구문제는 멀티미디어 요약(오디오/이미지 요약, 오디오 요약 및 이미지 요약)에 대하여 이용자들이 생각하는 각 요약에 대한 장점, 만족도 그리고 오디오 요약과 이미지 요약간의 상호 작용 효과는 무엇인가이다.

1.3 연구 방법

앞에서 기술한 세 가지 연구 문제들을 조사하기 위해서 선행 연구들과 국내외 디지털 도서관 또는 방송사의 영상물 관리 현황을 조사, 분석하여 오디오 요약, 이미지 요약 및 오디오/이미지 요약에 대한 알고리즘을 구성한다. 멀티미디어 요약을 구성하기 위하여 표본 비디오 10개를 선정한다. 각 표본 비디오의 오디오 내용과 프레임들을 분석한 후 본 연구에서 제안한 알고리즘을 이용하여 오디오 요약, 이미지 요약 및 오디오/이미지 요약을 구성하고 각 요약을 시청할 수 있는 세 종류의 실험 시스템들을 설계한다.

실험을 위해서 학부생(61명)과 대학원생(3명)으로 구성된 64명의 피조사자들을 이용한다. 64명을 세 개의 집단 즉, 그룹 1(22명), 그룹 2(21명) 및 그룹 3(21명)으로 구분한다. 그룹 1에게는 오디오/이미지 요약을 사용할 수 있는 실험 시스템 1을, 그룹 2에게는 오디오 요약을 사용할 수 있는 실험 시스템 2를 그리고 그룹 3에게는 이미지 요약을 사용할 수 있는 실험 시스템 3을 사용하도록 한다. 그런 다음 세 그룹의 실험 결과를 비교, 분석하기 위해서 SPSS 통계 패키지를 사용하여 다변량분석, 일원배치 분산분석, t 검증 등을 수행하였다.

1.4 연구 제한점

본 연구에서 설계한 오디오/이미지 요약은 오디오 요약을 먼저 구성한 후 오디오 요약문을 기준으로 그 오디오 정보가 포함되어 있는 지점에서 키프레임들을 추출하여 이미지 요약을 구성한 후 이 두 요약을 결합하여 구성하였다. 따라서 오디오/이미지 요약내의 오디오 및 이미지 정보가 항상 일치하지 않는 제한점이 있다.

2. 선행 연구

선행 연구에서는 텍스트 및 멀티미디어 정보의 차이점에 대하여 기술한 문헌, 비디오 검색 및 브라우징 환경에서 텍스트 대용물과 멀티미디어 대용물의 역할이 무엇인지 그리고 오디오 및 비주얼(이미지) 요약의 특성, 더 나아가 이들의 상호 작용 효과를 다루는 연구들을 살펴본다.

Paivio(1986)의 이중 부호론(dual-code theory)에 따르면 언어적 정보와 비언어적인 정보는 각기 다르게 처리되어 기억되며 비언어적인 정보(시청각 정보)는 언어적 정보 보다 더 오래 기억되는 경향이 있다. 많은 후속 연구들이 Paivio의 이론을 지지하였다. 예를 들어서, Ding et al.(1999)은 언어적 정보는 우리에게 객체에 대한 직접적이고, 논리적이면서 정확한 정보를 전달하는 대신 비주얼 정보는 풍부하면서 생생하고 구체적인 정보를 전달한다고 주장하였다.

이외에 Wildemuth et al.(2002)은 텍스트 대용물이 적합성을 결정하는 과정에서 도움을 줄 수 있는 반면, 멀티미디어 대용물은 효과적으로 텍스트 대용물을 확대시킨다고 기술하였다. Hughes et al.(2003)은 텍스트는 비디오가 무엇에 관한 것인지에 대한 정보를 전달하는 것인데 반해, 이미지는 비디오가 무엇인지에 관한 정보를 전달하는 것으로 특히 이미지 정보는 적합성 판단을 확실하게 할 때 자주 이용된다고 제안하였다. Yang과 Marchionini(2004)는 이용자들이 적합성 판정을 위해서 움직임이 포함된 이미지 요약을 좋아한다고 제안하였고, Iyer과 Lewis(2007)는 응답자의 75%가 이미지 요약이 비디오 자료에 대한 적합성 판정을 위해서 유용하게 사용될 수 있다고 하였다.

Song과 Marchionini(2007)는 오디오 및 이미지 요약을 결합한 멀티미디어 대용물의 효과가 오디오 요약 또는 이미지 요약만을 단독으로 사용했을 때 보다 더 효율적이라고 주장하고 있다. 또한 이 연구는 오디오 요약이 이미지 요약 보다 비디오 세그먼트에 대한 더 나은 이해를 가져다주지만 이용자들은 이미지 요약을

오디오 요약 보다 더 선호한다고 기술하였다. 아울러 이들 연구에서 이용자들은 오디오 요약을 통해서 비디오 내용을 확인하고 추가적인 가치를 얻는 것으로 나타났다.

Marchionini et al.(2009)은 비디오의 오디오 요약의 효율성을 테스트해 보기 위해서 25개의 비디오 자료와 48명의 피조사자들을 이용하였다. 이들의 연구 결과는 수작업으로 만들어진 오디오 요약이 비디오 내용을 가장 많이 전달하는 것으로 나타났다. 또한 자동으로 구성된 오디오 요약의 기능이 키워드와 비디오 빨리 감기(fast forwards)의 기능 보다 떨어지는 것으로 나타났다. 이외에 비주얼(이미지) 요소들은 이용자 경험에 주관적 가치를 부여하는 반면, 오디오 요약은 비디오 내용을 좀 더 명확하게 해 준다고 보고하고 있다. 이 연구는 또한 많은 이용자들이 오디오/이미지 요약이 비디오 내용을 더 정확하게 한다고 기술하고 있다. 그러나 24명 중 10명의 이용자들은 자동으로 구성된 오디오 요약과 이미지 요약이 결합된 경우에 오디오와 이미지 내용이 일치하지 않아서 오히려 혼란스러웠다고 답변하였다.

Kennedy et al.(2009)은 플리커 태그들로부터 위치와 이벤트 태그들을 추출하는 방법 그리고 태그와 위치로부터 추출된 지식을 이용하여 비주얼 특성을 분석하는 방안을 제안하였다. 이들은 비주얼 분석을 이용하여 자동으로 생성된 요약의 정확성이 비주얼 내용이 없는 경우와 비교하여 45% 이상 증가시켰다고 보고하고 있다.

Song et al.(2010)은 효율적인 비디오 요약 알고리즘을 생성하기 위해서 사람들이 비디오를 요약할 때 어떤 오디오 또는 비주얼 특성을 주목하는지에 대하여 실험을 수행하였다. 이들은 교육 다큐멘터리 비디오들을 표본 자료로 하여 실험을 수행한 결과, 비주얼 채널에서 비디오 요약을 위해서 오디오 정보가 없는 경우에는 텍스트, 수식 및 그래프가 중요시 되었고, 오디오 정보가 있는 경우에는 사람 얼굴, 자연 풍경 등이 자주 이용되었다. 오디오 채널에서는 한사람의 목소리와 자연 사운드가 선택되었고, 여러 사람들의 목소리들이 결합된 것은 일반적으로 선택되지 않았다.

김현희(2007)는 이미지 요약이 색인어 또는 요약문 추출 작업에 활용될 수 있을 뿐만 아니라, 디지털 도서관 환경에서 텍스트 초록과 같은 다른 메타데이터 요소들과 함께 사용된다면 이용자의 적합성 판정을 좀 더 용이하게 할 것이라고 제안하였다. 또한 이 연구는 이미지 요약이 이미지에 의존하여 내용을 전달하는 비디오인 경우 비디오의 의미 전달에 있어서 효과적이지만 만약 비디오가 주로 오디오를 통해서 정보를 전달하는 강의, 연설 자료인 경우 의미 전달의 효과가 낮아진다고 기술하였다.

이경미 외(2008)는 텍스타일 영상의 내용 데이터, 감성 데이터, 메타데이터를 결합시킨 영상 검색 시스템을 제안하였다. 제안된 시스템의 검색 절차는 메타데이터의 정보를 통해서 영상을 검색하게 된다. 검색된 영상 안에서 색상 히스토그램과 색상 스케치, 감성 히스토그램을 통하여 주어진 영상과 비슷한 영상들을 검색하게 된다. 이러한 방법은 검색 속도가 빨라짐으로써 검색 시간을 효율적으로 줄일 수 있으며, 또한 영상내용 뿐만 아니라 텍스트가 가지고 있는 의미를 보안하여 보다 효과적으로 검색을 할 수 있다고 제안하고 있다.

김현희(2009)는 비디오의 오디오 정보를 추출하여 자동으로 요약하는 알고리즘을 제안하고 좀 더 효율적인 오디오 요약을 구현하기 위해서는 오디오 정보의 특성에 맞춘 요약 기법에 대한 연구가 필요하다고 제안하고 있다. 특히 이 연구는 오디오 요약의 응집성이 떨어져 문맥이 자연스럽지 못한 점이 이용자의 만족도를 저하시키는 주요인이 된다고 언급하고 오디오 요약의 가독성을 향상시키기 위해서 이미지 요약을 오디오 요약과 함께 사용할 것을 권장하고 있다.

선행 연구를 종합해 볼 때, 비디오 검색 환경에 적용할 수 있는 비동시적으로 결합된 오디오/이미지 요약이 이용자 입장에서 비디오 의미 추출을 위해서 얼마나 효과적인지에 대한 체계적인 연구가 필요한 시점이 되었다고 생각한다.

3. 멀티미디어 요약 설계

3.1 개요

본 장에서는 세 개의 요약(오디오, 이미지, 오디오/이미지)을 알고리즘을 이용하여 설계하는데 요약 구성 과정은 오디오 요약과 이미지 요약을 만든 다음 이 두 개의 요약을 결합하여 오디오/이미지 요약을 구성하는 방식이기 때문에 오디오/이미지 요약의 설계에 대해서만 설명하고자 한다.

3.2 오디오/이미지 요약 구성

오디오/이미지 요약의 이미지와 음성(speech)의 연결성을 최대한 높이기 위해서 오디오 요약 알고리즘에 의해 자동으로 구성한 오디오 요약에서 핵심어로 언급되고 있는 인물이나 객체 등을 담고 있는 프레임들을 키프레임으로 수작업을 통해 선정하여 이미지 요약을 구성하였다. 즉, 오디오 요약을 먼저 구성하고 오디오 요약문을 기준으로 그 오디오 정보가 포함되어 있는 지점에서 키프레임을 추출하는 방식을 취하였다. 그런 다음, 오디오 요약과 이미지 요약을 결합하였다. 다음은 오디오 요약을 위한 알고리즘, 이미지 요약 방법 그리고 오디오/이미지 요약 구성 절차에 대해서 기술한다.

3.2.1 오디오 요약을 위한 알고리즘

오디오 요약 알고리즘을 구성하기 위해서 김현희(2009)가 제안한 각 문장의 가중치를 구하기 위해서 사용한 공식을 수정하여 활용하였다.[1] 수정된 공식(W)에서는 총 3개의 자질(위치, 표제어 및 태그)을 사용하였다. 비디오의 오디오 자료는 끝 부분 보다는 도입부 부분에서 비디오 주제에 대해 언급하는 경향이 있었다. 따라서 첫 6개 문장의 가중치는 '1'로 하였고 나머지는 모두 '0.8'로 처리하였다. 이외에 문장에 출현하는 표제어와 태그의 가중치는 각각 '1'과 '0.7'로 하였다.

$$\text{공식}(W) = \alpha + (\text{ti\_num} * 1) + (\text{tag\_num} * 0.7)$$

1) 본 연구에서 사용된 공식에서는 김현희(2009)가 제안한 방식에서 위치, 표제어 및 태그와 관련된 정보만을 사용하고 댓글 정보는 포함시키지 않았다.

W = 총 가중치
α = 문장 위치 가중치(첫 6개 문장의 가중치 -〉 1, 기타 -〉 0.8)
ti_num = 해당 문장에 포함된 비디오 표제어 수(비디오의 표제, 부제 및 소제목 등에 출현한 단어들 가운데 불용어를 제외한 단어수)
tag_num = 해당 문장에 포함된 비디오에 부여된 태그수

3.2.2 이미지 요약 방법

이미지 요약(스토리보드)을 구성하기 위해서 일반적으로 사용하는 비디오를 각 씬으로 나누고 각 씬에서 기계적으로 첫 번째나 마지막 프레임 또는 중간에 나타난 프레임을 키프레임으로 선택하는 대신 이미지와 오디오의 연결성을 높이는 방법을 채택하였다. 즉, 오디오 요약을 기준으로 그 오디오 정보가 포함되어 있는 지점에서 나타나는 프레임들 중에서 오디오 요약에서 핵심어로 언급되고 있는 인물이나 객체 등을 담고 있는 프레임들을 키프레임으로 선정하였다.

또한 이미지 요약의 반복 보기의 효과를 인정하여 각 표본 비디오에 대한 순차적 이미지 요약을 먼저 본 후 3단계 배열 모형(사건/객체, 인물, 배경)에 기초하여 구조화된 이미지 요약을 볼 수 있도록 두 종류의 이미지 요약을 설계하였다.

3.2.3 오디오/이미지 요약 구성 절차

첫째, 각 표본 비디오의 오디오는 스크립트를 이용하여 영문 텍스트로 변환하였다. 변환된 텍스트를 마침표나 물음표와 같은 구두점을 기준으로 하여 구분하였다. 그런 다음 각 문장에 순서대로 번호를 매겼다.

둘째, 각 문장에 위에서 기술한 공식(W)을 이용하여 가중치를 계산하여 할당하였다.

셋째, 문장의 가중치가 정해지면 가중치가 높은 순서대로 비디오의 재생시간에 따라서 4~10개의 문장을 선정하였다. 단 2단어 이하로 구성된 짧은 문장은 선정하지 않았다. 최종적으로 선택된 문장들을 텍스트에서 출현 순서대로 요약문을 구성한 후 한글로 번역하고 이를 음성 합성기인 매직 잉글리쉬 플러스를 이용하여 오디오 파일로 변환하여 오디오 요약을 구성하였다.

넷째, 이미지 요약은 오디오 요약에 포함된 스크립트를 분석하여 해당 스크립트가 말해지는 지점에서 핵심어로 언급되고 있는 인물이나 객체 등을 담고 있는 총 8개의 프레임들을 수작업으로 추출하여 구성한다.

다섯째, 오디오 및 이미지 요약을 결합한다.

앞에서 기술한 알고리즘, 요약 방법 및 요약 구성 절차에 의해서 만들어진 오디오/이미지 요약의 예는 〈그림 1〉에서 참고할 수 있다.

4. 오디오/이미지 요약의 오디오 및 이미지 정보간의 상호 작용 효과 분석

오디오/이미지 요약의 오디오 및 이미지 정보의 상호 작용 효과를 분석하고자 오디오/이미지 요약, 오디오 요약, 이미지 요약을 비디오의 의미 추출에 있어서의 정확도 그리고 이용자들이 인지하는 오디오 요약과 이미지 요약간의

상호작용 효과 그리고 각 요약에 대한 이용자들의 만족도를 비교, 분석하고자 한다. 다음은 실험 설계, 실험 결과 및 논의에 대해서 상술한다.

4.1 실험 설계

4.1.1 피조사자

실험을 위해서 학부생(61명)과 대학원생(3명)으로 구성된 64명을 활용하였다. 64명의 피조사자를 세 개의 집단, 즉 그룹 1(22명), 그룹 2(21명) 및 그룹 3(21명)으로 구분한다. 설문결과, 답변을 완전히 하지 않은 4명을 제외한 그룹 1(22명), 그룹 2(20명) 및 그룹 3(18명)을 포함한 총 60명의 자료를 분석하였다.

4.1.2 표본 비디오

표본 비디오는 유투브 사이트(www.youtube.com)에서 선정하였다. 멀티미디어 요약을 구성하기 위해서 표본 비디오 자료의 종류는 크게 두 가지로 선정하였다. 하나는 강의 및 연설 등 스크립트 중심의 비디오 유형 1(비디오 1-5번)과 또 다른 하나는 영상 중심의 비디오 유형 2(6-10번)로 구분하였다(〈표 1〉 참조). 비디오의 재생 시간은 5~11분 사이로 10개의 표본 비디오 중 5개는 10분 보다 길고 나머지 5개는 5~10분 사이이다. 영어 비디오들을 표본 비디오로 선정한 이유는 태그와 같은 소셜 메타데이터를 포함하고 있으면서 피조사자들에게 비디오 내용이 비교적 많이 알려지지 않았기 때문이다. 또한 본 연구에서 제안한 오디오/이미지 요약 알고리즘은 언어에 관계없이 모두 적용될 수 있기 때문이다.

4.1.3 실험 시스템 구현

각 그룹에게 이용하게 할 10개의 표본 비디오에 대한 세 개의 실험용 비디오 인터페이스들을 구현하였다. 즉, 각 비디오에 대한 세 가지 종류의 요약을 구현하였다. 그룹 1에게는 오디오/이미지 요약을 시청하게 한 후 각 비디오에 대한 요약문을 입력하게 하였다. 그룹 2와 그룹 3에게는 각각 오디오 요약과 이미지 요약(스토리보드)을 듣거나 보도록 한 후 그룹 1과 동일한 작업을 하도록 하였다. 〈그림 1〉은 그룹 1을 위한 순차적 이미지 요약과 오디오 요약을 동시에 시청하면서 비디오 1(Meet the Mentor)의 내용을 요약하여 입력하는 인터페이스이다. 〈그림 1〉의 오른쪽 위에 구조화된 스토리보드를 클릭하면 구조화된 이미지 요약을 볼 수 있다.

〈표 1〉 표본 비디오

번호	표제(재생시간(분:초))	번호	표제(재생시간(분:초))
1	Meet the Mentor(05:09)	6	How-to Make Mosaic Art(05:42)
2	Disability Services at ASU Libraries(10:05)	7	Thrive Food Storage- Spaghetti and Meatball(10:48)
3	Kate Lundy: What I do for Open Government(05:01)	8	Application and Preparation Limewash(06:27)
4	Learning English-Lesson Forty Three(Superstition)(10:41)	9	Experience The Learning Connexion(09:46)
5	President Obama at Michigan Commencement(10:00)	10	the Department of Dance at California State University, Long Beach(CSULB)(10:00)

〈그림 1〉 요약문 입력 화면(그룹 1)

요약문 입력이 끝나면 그 다음 해당 비디오에 대한 10개 항목(예, 미국 전력회사의 장애인 서비스에 대해 설명하고 있다)을 제시하고 해당 항목이 비디오 내용을 설명한 것이면 '맞다'를 체크하고 아니면 '틀리다'를 체크하도록 하는 화면이 나온다. 10개 비디오에 대한 요약문 입력과 항목 선택 작업이 모두 끝나면 마지막 단계에서 멀티미디어 요약에 대한 피조사자의 생각을 기술하도록 하였다. 즉, 그룹 1은 5개의 질문에 답변하도록 하였고, 그룹 2와 3은 2개의 질문에 각각 답변하도록 하였다.

4.1.4 실험 절차

실험은 컴퓨터실에서 실시되었다. 하나의 컴퓨터실을 세 구역으로 구분하여 각 그룹에게 해당 구역에 앉도록 하였다. 그룹 3(21명)을 제외한 그룹 1(22명)과 그룹 2(21명)는 헤드셋이나 이어폰을 사용하여 오디오 요약을 듣도록 하였다. 세 명의 실험 진행자가 각 그룹에게 10분 동안 실험 진행시 유의사항을 설명한 후 실험을 진행하였다.

피조사자들이 실험 시스템을 통해서 각 비디오에 대한 멀티미디어 요약을 시청하고 최소 세 문장 이상으로 요약문을 입력하게 하였다. 그 다음 단계로 적합한 항목(들)을 선택하도록 하였다. 이때 10개로 구성된 전체 항목을 보고 비디오 내용을 유추할 가능성이 높다고 판단하여 항목을 선택한 다음 이전 화면으로 돌아가 비디오 요약을 재작성하지 못하도록 제한하였다. 10개의 비디오에 대한 요약문 작성과 항목 선택이 끝나면 끝으로 이용자들이 인지하는 오디오 요약과 이미지 요약간의 상호작용 효과 그리고 각 요약에 대한 만족도를 기술하도록 하였다. 진행 시간은 유의 사항을 설명한 10분

을 포함하여 총 80분이 걸렸다.

4.1.5 비디오의 의미 파악 정확도 평가

비디오의 의미 파악 정확도 평가는 답변을 완전히 하지 않은 4명을 제외한 그룹 1(22명), 그룹 2(20명) 및 그룹 3(18명)을 포함한 총 60명에 의해서 작성한 600개의 사례를 분석하여 수행하였다. 각 비디오에 대한 요약문 정확도 평가는 2명의 연구원이 분석하였다. 점수 범위를 '0-20'로 하고 전혀 틀림(0점), 절반 미만 맞음(5점), 절반 맞음(10점), 절반 보다 더 맞음(15점) 그리고 완전히 맞음(20점)을 주도록 하였다. 두 사람의 점수 차이가 10점 미만인 경우에는 두 연구원의 매긴 점수들의 평균값을 정확도로 사용하였다. 그러나 만약 동일한 비디오에 대하여 두 연구원의 점수 차이가 10점 이상 차이가 나면 또 다른 연구원이 점수를 매기도록 하여 세 연구원이 매긴 점수들의 평균값을 정확도로 활용하였다.

항목 선택의 정확도는 각 비디오에 대한 10개 항목들의 점수 범위를 '0-20'으로 하고 전체가 맞으면 20점, 한 개 틀릴 때 마다 2점을 감점하는 방식으로 계산하였다. 예를 들어서 맞는 항목을 틀리다고 하든지 아니면 반대로 틀린 항목을 맞는다고 하든지 하면 2점을 감점하였다.

4.2 결과

4.2.1 비디오의 의미 파악 정확도

(1) 통계 분석 결과

① 다변량분석

본 연구에서는 독립변인의 주효과와 상호작용 효과를 검증하고 여러 종속변인의 선형조합에 대한 독립변인의 효과 검증이 가능한 다변량분석을 수행하였다. 여기서 독립 변인은 요약 유형(오디오/이미지 요약, 오디오 요약 및 이미지 요약), 비디오 유형(유형 1과 2) 그리고 이 독립변인들의 상호작용으로 구성되며, 종속변인은 요약문 정확도와 항목 선택 정확도가 된다. 분석 결과, 요약문 정확도와 항목 선택 정확도에서 모우 두 독립변인간의 상호작용 효과는 없고 요약 유형만이 두 개의 종속 변인에 영향을 미치는 것으로 나타났다(〈표 2〉 참조).

다음은 다변량분석의 사후 테스트 결과를 기초로 하여 요약 유형이 요약문 정확도 및 항목 선택 정확도에 미치는 영향을 상술하고자 한다.

(가) 요약 유형이 요약문 정확도에 미치는 영향

사후 테스트 결과에 의하면 요약문 정확도는 세 가지 요약 간에 통계적으로 유의미한 차이가 있는 것으로 나타났다(〈표 3〉과 〈표 4〉 참조).

〈표 2〉 다변량분석 결과

종속변인 / 독립변인	요약문 정확도 f(p)	항목 선택 정확도 f(p)
요약 유형	267.19(0.00)*	4.37(0.01)*
비디오 유형	1.20(0.27)	0.13(0.72)
요약 유형 * 비디오 유형	1.10(0.33)	0.48(0.62)

*는 0.05 수준에서 유의미함

〈표 3〉 요약문 정확도 평균(표준 편차)

요약문 정확도 \ 요약유형	오디오/이미지	오디오	이미지
비디오 유형 1	12.57(2.65)	11.49(2.69)	1.13(2.06)
비디오 유형 2	12.45(2.89)	11.48(2.82)	2.38(2.06)
전체(비디오 유형 1 & 2)	12.51(2.76)	11.48(2.75)	1.75(2.13)

〈표 4〉 사후 테스트 결과(요약문 정확도)

요약 유형(I)	요약 유형(J)	평균차(I-J)	표준 오차	유의확률
오디오/이미지	오디오	1.03	0.27	0.00
	이미지	10.76	0.47	0.00
오디오	오디오/이미지	-1.03	0.27	0.00
	이미지	9.73	0.47	0.00
이미지	오디오/이미지	-10.76	0.47	0.00
	오디오	-9.73	0.47	0.00

(나) 요약 유형이 항목 선택 정확도에 미치는 영향

사후 테스트 결과에 의하면 항목 선택의 정확도는 오디오/이미지 요약과 이미지 요약 그리고 오디오 요약과 이미지 요약간의 항목 선택의 정확도 차이가 통계적으로 유의미한 것으로 나타났으나 오디오/이미지 요약과 오디오 요약의 정확도간의 유의미한 차이는 없는 것으로 나타났다(〈표 5〉와 〈표 6〉 참조).

〈표 5〉 항목 선택 정확도 평균(표준 편차)

비디오 유형 \ 요약유형	오디오/이미지	오디오	이미지
비디오 유형 1	6.75(1.37)	6.67(1.40)	5.70(2.08)
비디오 유형 2	6.48(1.57)	6.43(1.40)	5.95(1.70)
전체(비디오 유형 1 & 2)	6.62(1.47)	6.55(1.41)	5.83(1.88)

〈표 6〉 사후 테스트 결과(항목 선택 정확도)

요약 유형(I)	요약 유형(J)	평균차(I-J)	표준 오차	유의확률
오디오/이미지	오디오	0.07	0.15	0.65
	이미지	0.79	0.26	0.00
오디오	오디오/이미지	-0.07	0.15	0.65
	이미지	0.72	0.26	0.00
이미지	오디오/이미지	-0.79	0.26	0.00
	오디오	-0.72	0.26	0.00

② 비디오 유형별 일원배치 분산분석과 사후 테스트 결과

비디오 유형별로 일원 배치 분산 분석과 사후 테스트를 수행한 결과에 의하면 요약문 정확도는 비디오 유형에 관계없이 세 가지 요약 간에 통계적으로 유의미한 차이가 있는 것으로 나타났다(〈표 3〉과 〈표 7〉 참조, 사후 테스트 결과는 생략함).

한편 항목 선택의 정확도는 비디오 유형에 따라서 결과값이 다르게 나왔다(〈표 5〉와 〈표 8〉 참조, 사후 테스트 결과는 생략함). 즉, 비디오 유형 1에서는 오디오/이미지 요약과 이미지 요약 그리고 오디오 요약과 이미지 요약간의 항목 선택의 정확도 차이가 통계적으로 유의미한 것으로 나타났으나 오디오/이미지 요약과 오디오 요약의 정확도간의 유의미한 차이는 없는 것으로 나타났다. 한편 비디오 유형 2에서는 세 요약간에 정확도 차이가 없었다.

(2) 가설 검증 결과

① 연구 가설 1과 2: 요약문 정확도는 세 가지 요약 간에 모두 통계적으로 유의미한 차이가 있는 것으로 나타났다(〈표 4〉 참조). 〈표 4〉에 나타난 것처럼 전체 비디오(비디오 유형 1 & 2)에서 오디오 요약, 오디오/이미지 요약 및 이미지 요약간의 비디오의 의미 추출의 정확도에 차이가 있을 것이라는 연구 가설 1은 검증되었다. 오디오/이미지 요약과 오디오 요약의 평균 정확도는 각각 12.51과 11.48로 나타나 큰 차이는 없게 나타났으나 통계적으로 유의미한 차이를 보였다. 이와 달리 이미지만을 본 경우는 평균 정확도가 1.75로 매우 낮게 나왔다.

〈표 7〉 일원배치 분산분석 결과(요약문 정확도)

	제곱합	df	평균제곱	F	유의확률
그룹간(비디오 유형 1)	2320.16	2	1160.08	147.68	0.00
그룹내(비디오 유형 1)	1783.14	227	7.85	-	-
전체(비디오 유형 1)	4103.30	229	-	-	-
그룹간(비디오 유형 2)	1816.98	2	908.49	108.40	0.00
그룹내(비디오 유형 2)	1902.49	227	8.38	-	-
전체(비디오 유형 2)	3719.47	229	-	-	-

〈표 8〉 일원배치 분산분석 결과(항목 선택 정확도)

	제곱합	df	평균제곱	F	유의확률
그룹간(비디오 유형 1)	17.30	2	8.65	3.80	0.02
그룹내(비디오 유형 1)	516.46	227	2.28	-	-
전체(비디오 유형 1)	533.76	229	-	-	-
그룹간(비디오 유형 2)	4.58	2	2.29	1.03	0.36
그룹내(비디오 유형 2)	506.82	227	2.23	-	-
전체(비디오 유형 2)	511.40	229	-	-	-

한편, 항목 선택의 정확도는 오디오/이미지 요약과 오디오 요약의 항목 선택의 정확도가 각각 6.62, 6.55로 나타났으며 이 두 정확도간에 통계적으로 유의미한 차이는 없었다. 다만 이미지의 항목 선택의 정확도(5.83)가 앞의 두 요약의 정확도들과 각각 통계적으로 유의미한 차이를 보였다. 이와 같이 세 요약간의 항목 선택의 정확도가 모두 경우에 유의미한 차이를 보이지 않아서 연구 가설 2는 검증되지 못했다(〈표 6〉 참조).

② 연구 가설 3과 4: 요약문 정확도는 비디오 유형 1에서 각각 오디오/이미지 요약의 정확도(12.57), 오디오 요약의 정확도(11.49) 및 이미지 요약의 정확도(1.13)간에 모두 통계적으로 유의미한 차이가 있는 것으로 나타났다(〈표 3〉과 〈표 7〉 참조). 따라서, 비디오 유형 1에서 오디오 요약, 오디오/이미지 요약 및 이미지 요약간의 비디오의 의미 추출의 정확도에 차이가 있을 것이라는 연구 가설 3은 검증되었다.

한편, 항목 선택의 정확도는 비디오 유형 1에서 오디오/이미지 요약의 정확도(6.75)와 이미지 요약의 정확도(5.70)간에 그리고 오디오 요약의 정확도(6.67)와 이미지 요약의 정확도(5.70)간에 차이가 있는 것으로 나타났으나 오디오/이미지 요약의 정확도(6.75)와 오디오 요약의 정확도(6.67)간에 차이가 없는 것으로 나타나 가설 4는 검증되지 못했다(〈표 5〉와 〈표 8〉 참조).

③ 연구 가설 5와 6: 요약문 정확도는 비디오 유형 2에서 각각 오디오/이미지 요약의 정확도(12.45), 오디오 요약의 정확도(11.48) 및 이미지 요약의 정확도(2.38)간에 모두 통계적으로 유의미한 차이가 있는 것으로 나타났다(〈표 3〉과 〈표 7〉 참조). 따라서, 비디오 유형 2에서 오디오 요약, 오디오/이미지 요약 및 이미지 요약간의 비디오의 의미 추출의 정확도에 차이가 있을 것이라는 연구 가설 5는 검증되었다. 항목 선택의 정확도는 비디오 유형 2에서 오디오/이미지, 오디오 및 이미지 요약의 정확도가 6.48, 6.43, 5.95로 각각 나타났으며 이들 요약간에 유의미한 차이가 없었다(〈표 5〉와 〈표 8〉 참조). 따라서 가설 6은 검증되지 못했다. 이는 이미지 요약이 이미지로 정보를 전달하는 비디오 유형 2에서 오디오/이미지 요약 또는 오디오 요약과 같은 효과를 냈다는 것을 보여준다.

4.2.2 이용자 관점

다음은 이용자에 대한 질문으로 그룹 1에게는 5개, 그룹 2와 3에게는 2개의 질문을 던져 각 질문에 대한 답변을 분석한 것이다.

(1) 오디오/이미지 요약을 시청한 그룹 1의 관점

① 질문 1: "오디오와 스토리보드(영상)가 함께 사용되어서 비디오 내용 파악에 어떤 영향을 주고 있습니까?"에 대한 질문에 "오디오와 이미지가 결합하여 비디오 의미 파악을 더 용이하게 한다(40.4%)"라는 답변과 "비디오에 따라 이 두 요약의 결합이 혼동을 주기도 하고 의미 파악을 더 용이하게 한다(40.4%)"라는 답변이 각각 가장 높게 나왔다(〈표 9〉 참조). 이는 대체로 오디오와 이미지가 결합하면 비디오 의미 파악을 더 용이하게 하지만 비디오 또는 피조사자에 따라서 이 두 요약의 결합이 오히려 혼동을 주는 경우도 있는 것으로 생각된다.

〈표 9〉 오디오/이미지 요약의 상호작용 효과(그룹 1)

항목	비율
오디오와 이미지가 결합하여 비디오 의미 파악을 더 용이하게 함	9(40.4%)
비디오에 따라 이 두 요약의 결합이 혼동을 주기도 하고 의미 파악을 더 용이하게 함	9(40.4%)
오디오와 이미지의 내용이 일치하지 않아서 혼동을 줌	2(9.1%)
기타(영상의 크기가 작아서 인지 어떤지 몰라도 오디오에 집중하게 됨, 영상의 내용은 크게 영향을 주지 않음)	2(9.1%)
합 계	22(100%)

혼동을 주는 가장 큰 이유는 오디오와 이미지의 내용이 일치하지 않은 점 때문으로 생각된다. 실제 몇몇 사례에서 오디오 요약의 정확도가 오디오/이미지 요약의 정확도 보다 더 높게 나온 경우가 있었다.

② 질문 2: "전반적으로 비디오 의미 파악을 위해서 더 유용하게 사용된 요약의 종류는 무엇입니까?"에 대한 질문에 "오디오 요약"이라는 답변이 77.3%로 가장 높게 나타났다. 그 다음으로 "오디오 및 영상 요약이 동일하게 유용하다"라는 답변이 18.2%, "이미지 요약"이 4.5%순으로 나타났다.

③ 질문 3: "오디오 요약이 좋은 점이 무엇입니까?"에 대한 질문에 "오디오 요약을 통해서 내용 파악이 용이하다"는 답변이 37.0%로 가장 많이 언급되었고 "오디오 요약을 통해서 구체적인 내용을 알 수 있다"는 답변이 33.3%로 그 다음으로 많이 언급되었다(〈표 10〉 참조).

④ 질문 4: "이미지 요약의 좋은 점이 무엇입니까?"에 대한 질문에 이미지 요약을 통해서 "비디오의 전체적인 흐름을 파악할 수 있다"는 답변이 34.6%로 가장 많이 언급되었다. 그 다음으로 "오디오 요약에서 들은 장면들을 실제로 볼 수 있어 내용 파악이 쉽다"가 19.2%, "오디오로는 표현하지 못하거나 말로 설명하기 애매한 내용들을 파악할 때는 이미지 요약이 더 확실한 것 같다"가 15.4%로 그 다음으로 많이 언급되었다. 이외에 이미지 요약이 시간의 구애가 없고, 오디오 요약을 다 듣는 시간 보다 짧아서 시간을 절약 할 수 있다는 답변이 각각 7.7%를 차지하였다(〈표 11〉 참조).

〈표 10〉 오디오 요약의 장점(그룹 1)

항목(비슷한 것끼리 묶음)(복수 응답)	비율
이미지 요약은 이미지만 보이기 때문에 대화 내용이나 내용 파악에 제한이 있다. 오디오 요약이 내용 파악에는 확실히 더 이미지 요약보다 좋으며 용이하다. 어떻게 전개될 것인지에 대한 전체적인 줄거리 파악이 용이하다.	10(37.0%)
이미지 요약만으로 알 수 없는 구체적인 내용에 대해서도 알 수 있다. 내용을 보다 구체적으로 유추할 수 있다.	9(33.3%)
말을 하는 것이 들리므로 바로 서술할 수 있다.	2(7.4%)
기 타	6(22.3%)
합 계	27(100%)

〈표 11〉 이미지 요약의 장점(그룹 1)

항목(비슷한 것끼리 묶음)(복수 응답)	비율
비디오의 전체적인 흐름을 파악할 수 있다. 이미지를 보고 대강의 내용을 추측할 수 있어 좋다.	9(34.6%)
오디오 요약에서 들은 장면들을 실제로 볼 수 있어 내용파악이 쉽다.	5(19.2%)
오디오로는 표현하지 못하거나 말로 설명하기 애매한 내용들을 파악할 때는 이미지 요약이 더 확실한 것 같다.	4(15.4%)
내용에 대한 이미지를 상상이 아닌 실제로 볼 수 있다.	2(7.7%)
정지된 화면이기 때문에 오디오 요약처럼 시간의 구애가 없다.	2(7.7%)
오디오 요약을 다 듣는 시간 보다 짧아서, 시간을 절약 할 수 있다.	2(7.7%)
기 타	2(7.7%)
합 계	26(100%)

⑤ 질문 5: “오디오/이미지 요약이 좋은 점이 무엇입니까?”에 대한 질문에 “하나의 요약으로 이해하기 힘든 내용을 오디오와 이미지를 병행하여 봄으로서 이해가 빠르다”라는 답변이 48.0%로 가장 높게 나왔다. 즉, 오디오를 바탕으로 구체적으로 내용을 유추하고, 이미지가 그 유추한 부분에 확신을 주거나 세밀한 수정을 가능하게 하여 짧은 시간에 대략적인 이해가 가능하다고 답변하였다. 그 다음으로 “오디오/이미지 요약이 오디오 및 이미지 요약의 단점들을 보완해준다”는 답변이 40.0%로 그 다음으로 높게 나타났다(〈표 12〉 참조).

(2) 오디오 요약을 청취한 그룹 2의 관점

① 질문 1: “오디오 요약이 비디오 내용 파악에 도움을 주었는가?”에 질문에 대한 답변으로 “보통이다”가 65%(13명)으로 가장 높게 나왔고, “도움을 주었다”가 35%(7명)으로 그 다음으로 높게 나타났다.

② 질문 2: “일반적으로 이미지 요약과 비교하여 오디오 요약이 좋은점으로 생각되는 것은 무엇인가?”에 대한 답변으로 “비디오의 주제를 쉽게 알 수 있다”라는 항목이 64.3%로 가장 높게 나왔다. 그 다음으로 “눈으로 볼 수 없는 사람에게 좋다”라는 항목이 14.3%로 그 다음으로 높게 나왔다(〈표 13〉 참조).

〈표 12〉 오디오/이미지 요약의 장점(그룹 1)

항목(비슷한 것끼리 묶음)(복수 응답)	비율
하나의 요약으로 이해하기 힘든 내용을 오디오와 영상을 병행하여 봄으로서 이해가 빠르다. 즉, 오디오를 바탕으로 구체적으로 내용을 유추하고, 영상이 그 유추한 부분에 확신을 주거나 세밀한 수정을 가능하게 하여 짧은 시간에 대략적인 이해가 가능하다. 또한 이미지 요약만으로는 파악할 수 없는 자세한 부분들을 오디오 요약을 통해 알 수 있다.	12(48.0%)
오디오와 이미지 요약의 단점들을 보완해준다.	10(40.0%)
기 타	3(12.0%)
합 계	25(100%)

〈표 13〉 오디오 요약의 좋은점(그룹 2)

항목(비슷한 것 끼리 묶음)(복수 응답)	비율
비디오의 주제를 쉽게 알 수 있다.	18(64.3%)
눈으로 볼 수 없는 사람에게 좋다.	4(14.3%)
시간이 절약된다.	2(7.1%)
집중이 잘 된다.	2(7.1%)
내용을 상상할 수 있다.	2(7.1%)
합 계	28(100%)

(3) 이미지 요약을 본 그룹 3의 관점

① 질문 1: “이미지 요약이 비디오 내용 파악에 도움을 주었는가?”에 질문에 대한 답변으로 “보통이다”와 “도움을 주지 않다” 각각 50%(9명)로 동일하게 나왔다.

② 질문 2: “일반적으로 오디오 요약과 비교하여 이미지 요약이 좋은점으로 생각되는 것은 무엇인가”에 대한 답변으로 “실제 자료를 볼 수 있다”라는 항목이 33.3%로 가장 높게 나왔다. “이미지로 내용 추측이 가능하다”라는 항목과 “집중이 용이하다”라는 항목이 각각 29.6%, 14.8%로 그 다음으로 높게 나타났다(〈표 14〉 참조).

4.3 논의

4.3.1 비디오의 의미 파악 정확도

첫째, 오디오/이미지 요약의 오디오와 이미지 정보가 서로 결합될 때 오디오 요약 또는 이미지 요약만 단독으로 사용했을 때 보다 어떤 상호 작용 효과를 가지고 있는지 살펴보기 위한 연구 가설 1의 검증 결과, 오디오/이미지 요약, 오디오 요약 및 이미지 요약의 요약문 정확도는 각각 12.51, 11.48, 1.75로 통계적으로 서로 유의미한 차이를 보였다. 즉, 오디오/이미지 요약의 정확도가 다른 두 개 요약의 정확도 보다 통계적으로 유의미하게 높게 나타나 오디오 및 이미지 요약 정보의 상호 작용 효과가 있는 것으로 나타났다.

이때 이미지만을 본 경우는 평균 요약문 정

〈표 14〉 이미지 요약의 장점(그룹 3)

항목(비슷한 것 끼리 묶음)(복수 응답)	비율
실제 자료를 볼 수 있다.	9(33.3%)
이미지로 내용 추측이 가능하다.	8(29.6%)
집중이 용이하다.	4(14.8%)
시간이 절약된다.	2(7.4%)
영상의 분위기를 알 수 있다.	2(7.4%)
용량이 적어 오디오 보다 많은 이미지를 올려서 볼 수 있다.	2(7.4%)
합 계	27(100%)

확도가 1.75로 이용자들이 다른 메타데이터 참조 없이 8개의 키프레임으로 구성된 이미지 요약만으로 비디오 의미를 파악하는 것이 매우 어려움을 보여주고 있다. 김현희(2007)의 연구와 비교하여 본 연구의 이미지 요약의 정확도가 상대적으로 낮게 나온 이유는 키프레임수와 비디오 내용의 난이도가 영향을 미친 요인일 가능성이 있다. 즉, 김현희 연구에서 사용된 키프레임수가 12~22개로 본 연구에서 사용한 8개 보다 많았다. 또한 김현희 연구에서 사용한 비디오는 주로 NASA 등에서 제공하는 중고등학생들을 대상으로 한 비디오였다면 본 연구에서 사용한 비디오는 유투브에서 추출한 다양한 주제를 다루고 있는 비디오이기 때문에 그만큼 내용이 어려울 수 있다는 점 때문이다.

한편, 연구 가설 2의 검증 결과, 항목 선택의 정확도에서는 예측과 달리 오디오/이미지 요약과 오디오 요약의 항목 선택의 정확도가 각각 6.62, 6.55로 나타났으며 이 두 정확도간에 통계적으로 유의미한 차이는 없었다. 다만 이미지의 항목 선택의 정확도(5.83)가 앞의 두 요약의 정확도들과 각각 통계적으로 유의미한 차이를 보였다. 따라서 항목 선택의 정확도에서는 오디오 및 이미지 요약 정보의 상호 작용 효과가 입증되지 못했다.

둘째, 비디오 유형에 따라서 이러한 상호 작용 효과가 어떻게 달라지는지 살펴보기 위한 연구 가설 3-6의 분석 결과, 요약문 정확도는 비디오 유형에 관계없이 세 가지 요약 간에 통계적으로 유의미한 차이가 있는 것으로 나타나 비디오 유형에 의해서 상호 작용 효과가 달라지지 않음을 확인하였다.

항목 선택의 정확도는 비디오 유형에 따라서 정확도가 다르게 나와 비디오 유형이 상호 작용 효과에 영향을 미치고 있음을 확인할 수 있었다. 즉, 비디오 유형 1에서는 오디오/이미지 요약과 이미지 요약 그리고 오디오 요약과 이미지 요약간의 항목 선택의 정확도 차이가 통계적으로 유의미한 것으로 나타났다. 그러나 오디오/이미지 요약과 오디오 요약의 정확도간의 유의미한 차이는 없는 것으로 나타났다. 이러한 결과는 오디오 요약도 오디오/이미지 요약과 비교하여 그 효율성이 크게 뒤지지 않음을 나타낸 것이다. 따라서 오디오 요약이 음성검색과 함께 스마트폰 환경에서 효율적으로 활용될 수 있을 것으로 생각된다.

한편 비디오 유형 2에서 세 요약간에 항목 선택의 정확도 차이가 없는 것으로 나타났다. 이는 이미지 요약의 효율성은 다른 두 개의 요약과 비교할 때 상대적으로 떨어지지만 영상 중심의 비디오 유형에서는 유용성이 높게 나오는 것을 확인할 수 있었다.

4.3.2 이용자 관점

오디오/이미지 요약에서 오디오와 이미지가 함께 사용되어서 비디오 내용 파악에 어떤 영향을 미치는 가에 대한 질문에 "오디오와 이미지가 결합하여 비디오 의미 파악을 더 용이하게 한다"와 "비디오에 따라 이 두 요약의 결합이 혼동을 주기도 하고 의미 파악을 더 용이하게 하기도 한다"는 답변이 각각 40.4%로 가장 높게 나왔다. 또한 오디오/이미지 요약이 좋은 점으로, "오디오와 이미지를 병행하여 봄으로서 이해가 빠르다"는 답변이 48.0%로 가장 높게 나왔다. 즉, 오디오를 바탕으로 구체적으로 내용을 유추하고, 이미지가 그 유추한 부분에 확

신을 주거나 세밀한 수정을 가능하게 하여 짧은 시간에 대략적인 이해가 가능하다고 하였다.

이러한 조사 결과는 대체로 오디오와 이미지 정보가 비동시적으로 결합되는 경우에도 이 두 정보가 상호작용하여 비디오 의미 파악을 더 용이하게 하지만 비디오 또는 피조사자에 따라서 이 두 요약의 결합이 오히려 혼동을 줄 수 있는 경우도 있는 것으로 나타났다.

혼동을 주는 가장 큰 이유는 오디오/이미지 요약의 경우 오디오와 이미지간의 비동시성으로 인하여 오디오와 이미지의 내용이 일치하지 않은 경우도 있기 때문으로 생각되며 비디오 또는 이용자에 따라서 어떤 일관성 있는 패턴을 보이는지 좀 더 세밀한 분석이 필요해 보인다. 다만 본 연구의 경우에는 그룹 1과 그룹 2 즉, 서로 다른 그룹을 비교하는 것이기 때문에 이용자 분석은 하지 않고 비디오에 따라서 어떤 차이를 보이는지 살펴보았다.

분석 결과, 실제 총 200개 사례 중 83개(42%)만이 오디오/이미지 요약의 요약 정확도가 오디오 요약의 정확도 보다 높게 나타났다. 반면에 66개 사례(33%)에서 오디오/이미지 요약과 오디오 요약간의 정확도 차이가 전혀 없는 것으로 나타났다. 더 나아가 50개 사례(25%)에서 오디오 요약의 정확도가 오디오/이미지 요약의 정확도 보다 더 높게 나왔다.

비디오별로 분석한 결과 〈표 15〉와 같은 결과가 나왔다. 10개의 비디오 중 오디오/이미지 요약의 정확도가 오디오 요약의 정확도 보다 통계적으로 유의미한 차이를 보이면서 높게 나오는 경우는 비디오 5번과 6번뿐이었다.

비디오 5는 오바마의 미시간 대학의 졸업식 연설 비디오로 오디오 요약에서 미시간이라는 단어가 나오고 졸업식 연설 장면을 보여주는 이미지 요약(예, 졸업식 가운 입은 오바마 얼굴)을 통해서 피조사자들이 어떤 비디오인지 파악이 대체로 쉬웠다고 생각한다. 특히 모든 오디오 요약은 여성의 목소리로 제작하였기 때문에 일부 피조사자들은 남녀 차이 때문에 혼란이 왔다고 답변하기도 하였다. 따라서 앞으로 오디오 요약을 제작할 때 비디오의 원래 목소리에 맞춰서 남성 또는 여성 목소리로 제작하는 방안을

〈표 15〉 비디오별 오디오/이미지 및 오디오 요약의 요약문 정확도간의 t검증 결과

비디오 번호	정확도(평균/표준편차)		t값	유의확률(p)
	오디오/이미지	오디오		
1	12.13(2.84)	10.88(2.72)	1.39	0.18
2	11.63(1.86)	11.00(2.05)	0.93	0.37
3	13.13(2.42)	12.50(2.56)	0.71	0.49
4	13.50(3.38)	12.63(3.49)	0.92	0.37
5	12.50(2.29)	10.38(1.68)	3.34	**0.00***
6	14.75(3.33)	12.50(2.81)	2.27	**0.04***
7	12.25(2.55)	11.25(3.29)	1.17	0.26
8	11.25(1.72)	10.38(2.60)	1.32	0.20
9	12.50(2.81)	11.38(2.36)	1.48	0.15
10	11.50(2.62)	11.88(2.80)	**-0.53**	0.60

고려해 볼 수 있다. 다음으로 비디오 6은 모자이크 예술을 만드는 방법에 대한 비디오로 비디오 유형 2에 속하기 때문에 예측한 대로 오디오/이미지 요약의 정확도 평균값이 오디오 요약의 정확도 평균값 보다 높게 나왔다.

한편 비디오 유형 2 범주에 속하면서 무용 학과에 대한 홍보를 다루는 비디오 10은 예측과 반대로 유의미한 차이는 없었지만 오디오 요약의 정확도가 오디오/이미지 요약의 정확도 보다 높게 나왔다.

이러한 분석 결과는 Marchionini 등(2009)의 연구에서 주장하는 즉, 오디오/이미지 요약 내의 이미지 및 오디오 정보가 일치하지 않아서 이용자들이 비디오 의미 파악을 하는데 있어서 별로 도움을 받고 있지 않음을 확인할 수 있었다. 따라서 이미지와 오디오간의 동시성을 기반으로 하여 구성한 비디오 스킴과 같은 비디오 요약은 이러한 문제점을 해결해 줄 것으로 생각되며, 앞으로 비디오 스킴이 디지털 도서관 검색 환경에서도 활용될 수 있기를 기대해 본다.

한편 오디오 요약의 좋은 점으로는 그룹 1의 경우 "내용 파악이 용이하다"는 답변이 37.0%로 가장 많이 언급되었고, "구체적인 내용을 알 수 있다"는 답변이 33.3%로 그 다음으로 많이 언급되었다. 그룹 2의 경우 "오디오 요약을 통해서 비디오의 주제를 쉽게 알 수 있다"는 답변이 64.3%로 가장 높게 나왔다. 이외에 이미지 요약의 좋은 점으로는 그룹 1의 경우 "비디오의 전체적인 흐름을 파악할 수 있다"는 답변이 34.6%로 가장 많이 언급되었다. 그 다음으로 "오디오 요약에서 들은 장면들을 실제로 볼 수 있어 내용 파악이 쉽다"는 답변이 19.2%로 나타났다. 그룹 3의 경우 "이미지 요약을 통해서 실제 자료를 볼 수 있는 점이 좋았다"는 답변이 33.3%로 가장 높게 나왔다. 이와 같이 오디오 요약과 비디오 요약의 특성과 역할이 서로 조금 다른 것으로 나타났다. 즉, 오디오 요약은 비디오의 구체적인 내용을 파악할 수 있게 한다는 답변이 많이 언급된 반면 이미지 요약은 비디오의 전체적인 흐름을 파악할 수 있어서 좋았다는 답변이 많이 언급되었다.

5. 결 론

본 연구는 오디오 요약과 이미지 요약이 비동시적으로 결합되어 구성된 오디오/이미지 요약이 만들어졌을 때 이 요약의 오디오 및 이미지 정보가 어떤 상호 작용 효과를 가지고 있는지, 또한 비디오 유형에 따라서 이러한 상호 작용 효과가 어떻게 달라지는지 조사하였다. 이를 위해서 오디오/이미지 요약, 오디오 요약 및 이미지 요약을 비디오의 의미 추출에 있어서의 정확도 즉, 요약문 및 항목 선택의 정확도와 이용자들의 이 세 가지 요약에 대한 관점을 서로 비교, 분석하였다. 분석 결과를 요약하면 다음과 같다.

첫째, 요약문 정확도에서는 오디오/이미지 요약에서 오디오와 이미지 정보의 상호 작용 효과를 확인하였다. 즉, 오디오/이미지 요약의 요약문 정확도(12.51)가 오디오 요약의 정확도(11.48) 및 이미지 요약의 정확도(1.75) 보다 더 높게 나타났다. 항목 선택의 정확도에서는 예측과 달리 오디오/이미지 요약과 오디오 요약의 항목 선택의 정확도가 각각 6.62, 6.55로 나타났으며 이 두 정확도간에 통계적으로 유의미한 차이는 없었다. 다만 이미지의 항목 선택의 정확

도(5.83)가 앞의 두 요약의 정확도들과 각각 통계적으로 유의미한 차이를 보였다. 따라서 항목 선택의 정확도에서는 상호 작용 효과가 입증되지 못했다.

둘째, 비디오 유형에 따라서 이러한 상호 작용 효과가 어떻게 달라지는지 살펴본 결과, 요약문 정확도는 비디오 유형에 관계없이 세 가지 요약 간에 통계적으로 유의미한 차이가 있는 것으로 나타나 비디오 유형에 의해서 상호 작용 효과가 달라지지 않음을 확인하였다. 항목 선택의 정확도는 비디오 유형에 따라서 정확도가 다르게 나와 비디오 유형이 상호 작용 효과에 영향을 미치고 있음을 확인할 수 있었다.

셋째, 이용자들의 멀티미디어 요약에 대한 관점을 조사한 결과는 다음과 같다. 이용자들은 오디오/이미지 요약은 오디오와 이미지를 병행하여 봄으로서 비디오 내용에 대한 이해를 빠르게 하고 또한 오디오를 바탕으로 구체적으로 내용을 유추하고, 이미지가 그 유추한 부분에 확신을 주거나 세밀한 수정을 가능하게 하여 짧은 시간에 대략적인 이해가 가능하다고 답변하였다. 한편 오디오/이미지 요약에서 오디오 및 이미지 정보의 결합이 혼동을 주기로 한다는 이용자의 답변이 있었고, 또한 실제 몇몇 사례에서 오디오 요약의 정확도가 오디오/이미지 요약의 정확도 보다 더 높게 나온 경우가 있었다. 이는 오디오/이미지 요약내의 오디오와 이미지 정보간의 비동시성으로 인하여 내용이 서로 일치하지 않은 경우도 있기 때문으로 생각된다.

본 연구 결과는 디지털 도서관 또는 모바일 환경에서 멀티미디어 요약을 구성하여 비디오 자료의 검색 및 브라우징의 효율성을 높이기 위한 방안을 모색하는데 활용될 수 있을 것이다.

참 고 문 헌

[1] 김현희. 2007. 비디오 자료의 의미 추출을 위한 영상 초록의 효용성에 관한 실험적 연구. 『정보관리학회지』, 24(4): 53-72.

[2] 김현희. 2009. 비디오의 오디오 정보 요약 기법에 관한 연구. 『정보관리학회지』, 26(3): 169-188.

[3] 이경미 외. 2008. 내용, 감성, 메타데이터의 결합을 이용한 텍스타일 영상 검색. 『한국인터넷정보학회논문집』, 9(5): 99-108.

[4] Ding, W., et al. 1999. "Multimodal surrogates for video browsing." *Proceedings of the fourth ACM Conference on Digital Libraries*, 85-93. Berkeley, CA.

[5] Gunther, R., Kazman, R., & MaccGregor, C. 2004. "Using 3D sound as a navigational aid in virtual environments." *Behaviour and Information Technology*, 23(6): 435-446.

[6] Hughes, A., et al. 2003. "Text or pictures? an eye-tracking study of how people view digital video surrogates." *Proceedings of CIVR 2003*, 271-280.

[7] Iyer, H., & Lewis, C. 2007. "Prioritization strategies for video storyboard keyframes." *Journal of American Society for Information Science and Technology*, 58(5): 629-644.

[8] Kennedy, L., Naaman, M., Ahern, S., Nair, R., & Rattenbury, T. 2007. "How Flickr helps us make sense of the world: Context and content in community-contributed media collections." *Proceedings of ACM Multimedia, 2007*. Augsburg, Germany. [online]. [cited 2010.5.10]. 〈http://infolab.stanford.edu/~mor/research/kennedyMM07.pdf〉.

[9] Kristin, B., et al. 2006. *Audio Surrogation for Digital Video: A Design Framework*. UNC School of Information and Library Science(SILS) Technical Report TR 2006-21.

[10] Marchionini, G., et al. 2009. "Multimedia surrogates for video gisting: Toward combining spoken words and imagery." *Information Processing and Management*, 45: 615-630.

[11] Paivio, A. 1986. *Mental Representations*. New York: Oxford University Press.

[12] Schmandt, C., & Mullins, A. 1995. "AudioStreamer: Exploiting simultaneity for listening." *CHI 95 Conference Companion 1995*, 218-219.

[13] Song, Y., & Marchionini, G. 2007. "Effects of audio and visual surrogates for making sense of digital video." *Proceedings of CHI 2007*, 867-876. San Jose, CA, USA.

[14] Song, Y., Marchionini, G., & Oh, C. 2010. "What are the most eye-catching and ear-catching features in the video?: implications for video summarization." *Proceedings of the 19th International Conference on World Wide Web 2010*. Raleigh, North Carolina.

[15] Wildemuth, B., et al. 2002. "Alternative surrogates for video objects in a digital library: Users' perspectives on their relative usability." *Proceedings of the 6th European Conference on Digital Libraries*, 493-507. New York: Springer.

[16] Yang, M., 2005. *An Exploration of Users' Video Relevance Criteria*. Ph.D. diss., University of North Carolina at Chapel Hill.

[17] Yang, M., & Marchionini, G. 2004. "Exploring users' video relevance criteria: A pilot study." *Proceedings of the Annual Meeting of the American Society of Information Science and Technology*, Nov. 12-17, 2004. 229-238. Providence, RI.

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

[1] Kim, Hyun-Hee. 2007. "An experimental study on the effectiveness of storyboard surrogates in the meanings extraction of digital video." *Journal of the Korean Society for Information Management*, 24(4): 53-72.

[2] Kim, Hyun-Hee. 2009. "Investigating the efficient method for constructing audio surrogates of digital video data." *Journal of the Korean Society for Information Management*, 26(3): 169-188.

[3] Lee, Kyoung-Mi, et al. 2008. "Textile image retrieval integrating contents, emotion and metadata." *Journal of Korean Society for Internet Information*, 9(5): 99-108.