

뇌파정보를 활용한 영상물 요약 알고리즘 설계와 평가*

Design and Evaluation of Video Summarization Algorithm based on EEG Information

김 현 희 (Hyun-Hee Kim)**

김 용 호 (Yong-Ho Kim)***

목 차

- | | |
|-----------|--------------|
| 1. 서 론 | 4. 비디오 스킵 구성 |
| 2. 이론적 배경 | 5. 비디오 스킵 평가 |
| 3. 실험 설계 | 6. 결 론 |

초 록

본 연구는 비디오 스킵의 자동 생성을 위한 비디오 요약 알고리즘을 제안하고 이를 평가하였다. 제안된 알고리즘은 ERP(Event Related Potentials) 기반의 주제 적합성 모형, MMR(Maximal Marginal Relevance) 기법 및 판별분석기법을 사용하여 구현하였다. 제안한 ERP/MMR 기반 알고리즘을 이용하여 구성된 비디오 스킵의 품질과 유용성을 내재적 및 외재적 평가를 통해서 검증하였다. 내재적 및 외재적 평가에서 ERP/MMR 방법들의 평가 점수들은 각각 경쟁 기준으로 사용한 SBD(Shot Boundary Detection) 방법의 평가 점수 보다 유의미한 차이를 보이며 높게 나왔다. 그러나 이 두 평가에서 ERP/MMR($\lambda=0.6$) 방법의 평가 점수와 ERP/MMR($\lambda=1.0$) 방법의 평가 점수 간에 통계적으로 유의미한 차이는 없는 것으로 나타났다.

ABSTRACT

We proposed a video summarization algorithm based on an ERP (Event Related Potentials)-based topic relevance model, a MMR (Maximal Marginal Relevance), and discriminant analysis to generate a semantically meaningful video skim. We then conducted implicit and explicit evaluations to evaluate our proposed ERP/MMR-based method. The results showed that in the implicit and explicit evaluations, the average scores of the ERP / MMR methods were statistically higher than the average score of the SBD (Shot Boundary Detection) method used as a competitive baseline, respectively. However, there was no statistically significant difference between the average score of ERP/MMR ($\lambda = 0.6$) method and that of ERP/MMR ($\lambda = 1.0$) method in both assessments.

키워드: 사건관련유발전위, ERP/MMR 모형, 주제 적합성 모형, 비디오 스킵, 내재적 평가, 외재적 평가
ERP, ERP/MMR Model, Topic Relevance Model, Video Skim, Implicit Evaluation,
Explicit Evaluation

* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A5A2A01023839).

** 명지대학교 문헌정보학과 명예 교수(kimhh@mju.ac.kr) (제1저자)

*** 부경대학교 신문방송학과 교수(kimyho@pknu.ac.kr) (교신저자)

논문접수일자: 2018년 10월 16일 최초심사일자: 2018년 10월 16일 게재확정일자: 2018년 11월 13일
한국문헌정보학회지, 52(4): 91-110, 2018. [http://dx.doi.org/10.4275/KSLIS.2018.52.4.091]

1. 서론

1.1 연구의 필요성과 목적

동영상 자료의 급속한 증가로 인하여 많은 연구들이 이를 효율적으로 검색할 수 있는 상호작용적 인터페이스(human computer interface, HCI) 설계를 중심으로 진행되고 있다. 이러한 시스템 설계와 활용에서 가장 어려운 문제 중 하나는 메타데이터(예, 영상물 요약)에 의해서 파악한 영상물의 주제와 시청자들이 그 영상물의 전체 내용을 시청한 후 기억하고 추론하는 주제 간에 의미 격차(semantic gap)가 발생한다는 점이다.

영상물 요약은 영상물 자료를 검색, 색인 및 브라우징하는 과정에서 이용자들이 검색시스템을 통하여 필요한 영상물을 효율적으로 이용하는 데 매우 중요하게 활용될 수 있는 메타데이터 요소이다. 영상물 자체의 정보는 구체적으로 화면의 색과 구성, 인물이나 객체의 형태와 움직임, 영상물의 구성 단위인 하나의 연속적인 화면인 쇼트(shot)나 장면(scene)의 흐름과 단절, 그리고 영상효과를 의미한다. 비디오 요약을 위한 키프레임 또는 키쇼트의 추출 방식들은 대부분 쇼트나 장면 단위의 세그먼트로 비디오를 구분한 후 이러한 세그먼트들에서 정해진 알고리즘을 이용하여 키프레임 또는 키쇼트를 추출하는 방법들이 있다(Chen, Delannay and Vleeschouwer 2011; Mishra et al. 2015). 이외에 비디오에 나타나는 인물이나 객체와 같은 내용 특징이 비디오 요약의 기준으로 사용되기도 한다(Tavassolipour, Karimian and Kasaei 2014). 이와 같이 쇼트 또는 장면의 경계 검출

(boundary detection) 방법과 내용 특징 추출 알고리즘의 급속한 향상에도 불구하고, 의미 격차 문제로 인하여 효과적인 비디오 요약 기법을 제시하지 못하고 있다는 비판도 있다(Hu et al. 2011).

이러한 의미 격차를 줄이기 위해서 Mehmood 외(2016)는 특정 프레임이 시각적 주목 산출 모형(visual attention model)에 의해서 시각적으로 시청자들에게 중요하다고 생각되면 이를 키프레임으로 추출하는 방안을 제안하였다. 또한 IBM사의 왓슨(Watson)은 상업영화 “The Morgan”의 트레일러를 자동구성하기 위해서 해당 영화와 같은 장르의 영화 100여편과 그 예고편들을 훈련 데이터(training data)로 사용하여 영화와 예고편간의 관계를 딥러닝 방식으로 분석하였다. 분석 결과, 감성반응이 중요한 변수였다는 점을 밝혀냈고 이를 기반으로 하여 예고편을 제작하였다(IBM Research Corp 2016). IBM사가 기준으로 사용한 감성 반응은 관객의 감정 반응이라기 보다는 감각 정보에 대한 즉각적인 반응에 더 가까워 보인다.

이와 같이 시각적인 자극에 대하여 사람이 주목하는 것은 거의 반사적인 반응으로 보이고 자극의 형식적인 특성을 부호화하는 지각도 거의 반사적인 반응으로 볼 수 있다. 따라서, 의미 격차의 근본적인 문제는 이러한 반사적인 주목과 지각 단계와 연결되어 있지 않아 보인다. 이는 시청자가 들어오는 외부 콘텐츠(자극)에 어떻게 반응하는지 즉 자극의 시맨틱 의미를 어떻게 숙고하고 이러한 의미를 어떻게 일관된 심적 표상(mental representation)으로 통합하는지를 파악하는 과정과 관련되어 있는 것으로 보인다. 바로 이 점 때문에 시청자들이 동영상 자료의 주제를 어떻게 이해하는지를 파악하는 것

이 매우 중요한 문제가 된다.

최근 연구들은 주제 적합성 판단 과정을 이해하기 위해서 신경과학적 연구 방법들을 이용하기 시작하였다(Mostafa and Gwizdka 2016). 예를 들어서, Sitnikova 외(2008)은 직접 인식, 감각 또는 운동 자극에 대한 뇌 반응으로 정의되는 사건관련유발전위(event related potentials, ERP)의 구성요소들을 이용하여 시청자의 비디오 주제 분석 과정을 기술하고자 하였다. 이를 위해서 Sitnikova et al.은 시청자에게 사운드만 있는 비디오 클립을 시청하게 하였다. 이때 비디오의 마지막 장면으로 비디오의 주제와 맞는 그리고 맞지 않은 쇼트를 각각 제시했을 때 시청자의 뇌파 반응에서 어떤 차이가 나타나는지 분석하여 주제 분석 과정에 대한 정보를 얻고자 하였다. 그러나 이들의 연구는 사운드를 제거한 동영상 실험 데이터로 사용했기 때문에 제한점이 있다고 할 수 있다. 김용호와 김현희(2016)는 실시간으로 비디오를 보면서 ERP의 구성 요소들을 이용하여 비디오에서 주제 관련 쇼트들을 추출하는 방법을 제안하였다.

본 연구는 비디오의 주제를 유추할 수 있으면서 주제 중복을 최소화시키는 비디오의 핵심 쇼트들로 구성된 비디오 스킴을 자동으로 생성하기 위한 비디오 요약 알고리즘을 제안하고 평가해 보았다. 제안된 알고리즘은 ERP-기반 주제 적합성 모형, 주제 중복을 최소화시킬 수 있는 최대 한계 관련성(maximal marginal relevance, MMR) 기법 및 판별분석기법을 사용하여 구현하였다. 이후 제안한 ERP/MMR 기반 알고리즘을 이용하여 구성된 비디오 스킴들의 품질과 유용성을 내재적 및 외재적 평가를 통해서

검증해 보았다.

1.2 연구의 방법

Mayer(2009)의 멀티미디어 학습 모형과 뇌파에 관한 선행 연구의 결과에 기반하여 비디오 스킴의 구성을 위한 ERP/MMR 기반 알고리즘에 대한 이론적인 틀을 수립하였다. 이후 M대학 학부생과 대학원생으로 구성된 피험자 23명과 4개의 다큐 비디오들을 실험 데이터로 이용하여 뇌파 실험을 수행한 후 알고리즘을 설계하였다. 이때 뇌파 데이터를 분석하기 위해서 반복측정(repeated measures) t-검증, 인공신경망(artificial neural networks, ANNs) 분석 및 판별분석을 수행하였다.

제안된 알고리즘을 평가하기 위해서 이를 이용하여 비디오 스킴을 구성한 후 비디오 스킴의 품질과 유용성을 내재적 평가와 외재적 평가를 통해서 측정하였다(정영미 2012). 본 연구에서 사용한 4개의 비디오와 이전 연구(김용호, 김현희 2016)에서 사용한 3개의 비디오를 합하여 총 일곱 개의 비디오들이 평가를 위해서 사용되었다.

구체적으로 내재적 평가에서는 제안한 알고리즘에 의해서 구성된 비디오 스킴을 표준 비디오 스킴과 ROUGE-1과 다양성 기준에 의하여 비교 평가해 보았다. 한편, 외재적 평가에서는 제안된 알고리즘에 의해서 구성된 비디오 스킴이 정보 브라우징에서 얼마나 효율적으로 활용되었는지를 측정해 보았다. 이를 위해서 제안된 알고리즘과 SBD 방식에 의해서 구성하는 비디오 스킴들을 39명의 피조사자들(세 그룹으로 구분)에게 시청하게 한 후 이들이 세 가지 유형의

비디오 스킴을 통해서 비디오 주제를 얼마나 정확하게 유추해 냈는지를 측정하였다. 이때 사용된 방법은 피조사자들에게 비디오 스킴을 시청하게 한 후 각 비디오에 대한 유추된 제목과 요약문을 기술하도록 하였다. 평가 결과를 분석하기 위해서 비모수적 방법인 윌콕슨 부호순위 검정(Wilcoxon matched-pairs signed-rank test)과 분산분석/사후 검증(LSD)을 수행하였다. 뇌파 데이터 분석과 내재적 및 외재적 평가에 사용된 통계 분석들은 모두 SPSS 23 통계 패키지를 이용하여 수행하였다.

1.3 용어 정의

본 연구에서 사용한 용어를 다음과 같이 정의하였다.

- 쇼트(shot): 영상으로 만들 때, 중간에 끊지 않고 촬영한 하나의 연속적인 화면이다.
- 신(scene): 동일 장소, 동일 시간 내에서 이루어지는 일련의 액션이나 대사로 보통 몇 개의 쇼트들로 구성된다.
- 비디오 스킴(video skim): 비디오의 핵심 쇼트(또는 신)들로 구성된 비디오 요약으로 영화의 예고편(trailer)과 유사하다.
- 사건관련유발전위(event related potentials, ERP): 외부의 자극 혹은 내부의 심리적 과정과 관련되어 뇌에서 발생하는 전위(potential)의 변화 즉 전위차를 나타낸다.
- N400: 자극 이후 400ms 근처에서 음전위(negative)로 나타나며 의미적 처리 과정과 관련이 있다고 알려져 있는 ERP의 구성요소이다(권준수 2000).

- P600: 자극 이후 600ms 근처에서 양전위(positive)로 나타나며 재구성, 회상 과정(recollective process) 또는 장기 기억에 저장되어 있는 정보를 근거로 하여 자극을 더 정교하게 처리하는 과정과 관련이 있다고 알려져 있는 ERP의 구성요소이다(권준수 2000).
- 기저선(baseline): 어떤 값을 측정하거나 평가할 때 사용하는 기본적인 값으로 실험에서 독립변인의 처치 효과를 측정하기 위해 처치를 가한 상태의 조건과 비교하기 위해 사용하는 일상적 행동을 지칭한다. 본 연구에서는 자극 제시 전 블랙 스크린이 제시되는 200ms 동안의 평균 진폭을 기저선으로 삼았다.
- 에폭(epoch): 뇌파의 분석 구간으로 본 연구에서는 자극 제시 전 200ms부터 자극 제시 후 1,000ms로 설정하여 각 에폭의 길이는 1.2초이다.

2. 이론적 배경

2.1 주제 적합성 모형

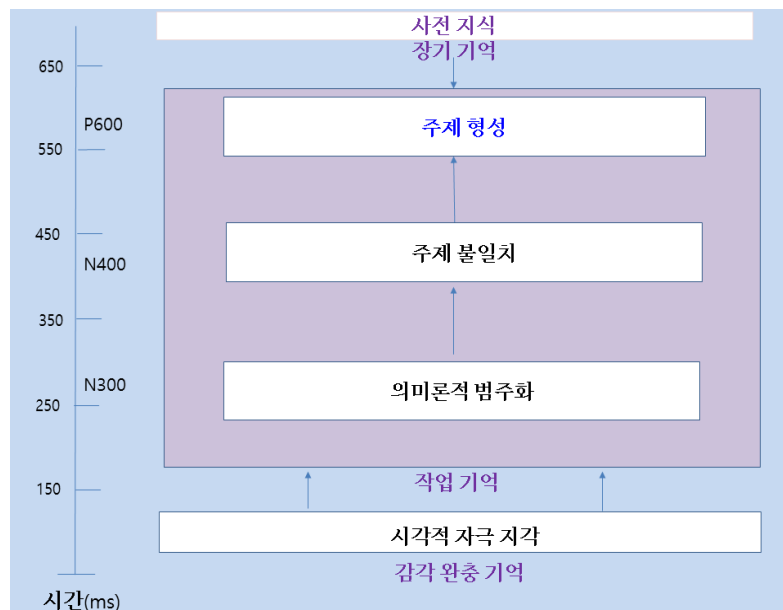
주제 적합성 모형은 시청자가 비디오를 시청할 때 비디오의 주제를 파악해 가는 동적 과정을 기술한다. 주제 판단 과정의 동적 특성을 이해하려면 시청자가 동영상 주제를 이해하는 과정을 검토해야 하는데 이를 위해서 Mayer(2009)의 멀티미디어 학습 모형을 분석해 보았다. Mayer 모형은 이중 채널(dual channels)과 능동적 처리(active processing)에 기반하는데 이중 채널

의 개념은 사람들은 정보를 처리할 때 두 개의 채널, 즉 언어 정보와 이미지 정보를 처리하는 두 개의 서로 다른 시스템을 갖고 있다고 가정한다. 더 나아가 Mayer는 사람들은 적합한 정보를 일관적 심적 표상으로 조직하고 이러한 표상을 장기기억장치에 있는 사전 지식과 통합하는 능동적 처리 능력을 가지고 있다고 보았다. 이러한 Mayer 모형을 ERP 요소를 이용하여 텍스트, 정지 이미지 및 동영상의 자극에 반응하는 시청자의 뇌파 반응을 기술한 선행연구들에 접목시켜 주제 적합성 모형(〈그림 1〉 참조)을 구성하였다.

비디오 주제를 결정할 때 두 가지 인지 과정 즉, 시각과 통합이 필요하다(Zhu et al. 2007). 여기서 통합은 다시 주제와 관련 없는 쇼트에 대한 의미론적 불일치 과정과 주제 관련 쇼트에 대한 문맥 갱신 과정의 두 단계로 구성된다고 볼

수 있다. 이러한 인지 과정을 고려해 볼 때 비디오 쇼트에 대한 시청자의 ERP 반응 측면에서 다음과 같은 몇 가지 논리적인 결과를 유추해 볼 수 있다.

첫째, 시청자가 시각적 자극을 감지하면 시청자는 이 자극을 분류하여 의미론적 범주화와 관련된 N300 효과(자극 이후 300ms 근처에서 음전위 최소값이 나타난다는 이론)를 나타낼 것이다(Mudrik et al. 2014). 이후 시청자가 비디오 쇼트를 주제와 무관한 것으로 인식하면 N400 효과를 보여줄 것이다. 이 효과는 주제와 무관한 비디오 쇼트에 대한 음전위 최소값이 주제 관련 비디오 쇼트에 대한 음전위 최소값 보다 더 커지는 것을 의미한다. 이와 같이 N400은 의미가 일치하지 않은 외부 자극에 반응하는 ERP 구성 요소로 알려져 있다(Sitnikova et al. 2008; 김용호, 김현희 2016).



〈그림 1〉 주제 적합성 모형

둘째, 시청자가 비디오 쇼트를 주제와 관련된 것으로 인식하면 P600 효과(자극 이후 600ms 근처에서 양전위 최대값이 나타난다는 이론)를 보여줄 것으로 가정하였다. P600은 후기 양전위 구성 요소(late positive component, LPC)로 지칭되기도 한다. P600은 외부 자극이 잠정적인 주제를 나타내는 기존의 개념들과 더 나아가 장기기억장치의 사전 지식과 통합하면서 주제를 갱신하는 문맥-갱신 단계(context-updating step)를 기술하는 것으로 알려져 있다(Burmester, Spalek, and Wartenburger 2014). 이외에 P600은 담화 구조의 유지 및 갱신(Schumacher and Hung 2012), 주제 전환(Xu and Zhou 2016), 적합성(Allegretti et al. 2015; Eugster et al. 2016), 담화의 내적인 재조직과 통합(Wang and Schumacher 2013) 등과 관련되어 있다고 보고되고 있다.

본 연구는 N400과 P600에 대한 두 개의 가설들을 수립하여 검증해 보고자 하였다. N300 효과는 주제 적합성에 관계없이 모든 비디오 쇼트에 나타날 것으로 예측되어 이에 대한 가설은 수립하지 않았다.

- 연구가설 1 (H1): 주제와 무관한 쇼트는 N400 효과를 나타낼 것이다.
- 연구가설 2 (H2): 주제와 관련된 쇼트는 P600 효과를 나타낼 것이다.

2.2 ERP/MMR 모형

ERP 기반 주제 적합성 모형을 MMR 기법과 결합하였다(Carbonell and Goldstein 1998). MMR 기법은 가장 적합성이 높은 쇼트를 선택하면서 동시에 이미 선택된 비디오 쇼트와의 중복

을 최소화하는 하기 위한 방법이다. 예를 들어서 비디오 스킴 구성 과정에서 이미 선택된 비디오 쇼트와 유사한 비디오 쇼트일수록 중요도를 낮춰서, 유사한 비디오 쇼트가 비디오 스킴의 구성 요소로 선택되지 않도록 하는 방법이다. MMR 기법에서 두 쇼트간의 유사도를 측정하기 위해서 프레임/용어 벡터 데이터를 이용하였다. 이에 따라서 ERP/MMR 모형은 비디오 주제에 가장 적합한 비디오 쇼트이면서 동시에 이미 선정된 비디오 쇼트간의 유사도를 최소화하는 방식으로 주요 쇼트들을 선택하도록 다음과 같이 구성되었다.

$$ERP/MMR = \arg \max_{S_i \in R \setminus S} [\lambda Dvalue(S_i) - (1 - \lambda) \max_{S_j \in S} Sim(S_i, S_j)],$$

ERP/MMR 모형에서 Dvalue(S_i)는 비디오 쇼트 S_i 의 주제 적합성을 나타낸 평균 판별점수이며 이 판별점수는 판별분석 결과로 얻는 판별함수($D = 0.102 \cdot F8_N400 + 1.145$)를 이용하여 산출한다(자세한 사항은 “3.4.2 판별 분석”에서 참조). R는 비디오 쇼트 집합에 있는 순위화된 비디오 쇼트, S는 비디오 스킴을 위해서 이미 선정된 비디오 쇼트들의 목록이다. 또한 $R \setminus S$ 는 S를 제외한 R 목록내에 있는 아직 선정되지 않은 비디오 쇼트들의 목록이다. 그리고 Sim(S_i, S_j)은 비디오 쇼트 S_i 와 S_j 간의 유사도이다. 본 연구에서는 이러한 유사도를 구하기 위해서 이미지 비교 프로그램(<https://sourceforge.net/projects/imagecomp/>)을 사용하여 두 비디오 쇼트간에 동일하거나 유사한 프

레이프의 수를 계산하여 측정하였다. 이외에 얼마나 많은 동일한 또는 유사한 단어가 비디오 쇼트의 오디오 정보에 포함되었는지 계산하여 오디오 내용의 유사성을 측정하였다.

적합성을 강조하거나 중복을 피하기 위해 결합 점수를 조정하는 데 사용되는 매개 변수인 λ (λ)는 0에서 1사이의 값을 갖는데 $\lambda = 1$ 인 경우에는 비디오 쇼트의 순위화는 판별 점수에만 기반하게 된다. 반대로 $\lambda = 0$ 인 경우에는 비디오 쇼트간의 최대 다양성을 반영한 순위화가 이루어진다. 본 연구에서는 테스트 결과, λ 값으로 0.6이 가장 적절하게 나타나 이를 채택하였다. 가장 높은 ERP/MMR 점수를 갖는 비디오 쇼트들이 반복적으로 선택되어 비디오 스킴을 구성한다. 비디오 요약에서 널리 사용되는 쇼트 경계 검출(shot boundary detection, SBD) 방법을 경쟁 기준(competitive baseline)으로 채택하여 ERP/MMR 모형에 관한 두 가지 가설을 수립하였다.

- 연구가설 3 (H3): ERP/MMR 모형에 의해서 구성된 비디오 스킴은 SBD 방식에 의해서 구성된 비디오 스킴 보다 품질과 유용성 측면에서 더 우수할 것이다.
- 연구가설 4 (H4): 중복을 최소화하는 ERP/MMR($\lambda = 0.6$) 모형에 의해서 구성된 비디오 스킴은 중복은 고려하지 않고 ERP

정보만 고려한 ERP/MMR($\lambda = 1.0$) 모형에 의해서 구성된 비디오 스킴 보다 품질과 유용성 측면에서 더 우수할 것이다.

3. 실험 설계

실험을 수행하기 위해서 피험자들을 모집하고 실험 재료들을 선정한 후 실험 절차를 설계하였다. 설계된 절차에 따라 실험을 수행한 후에 수집한 뇌파 정보를 분석하여 연구 가설 1과 2를 검증해 보았다. 실험을 수행하기 전에 명지대학교 기관생명윤리위원회의 승인을 받았다 (IRB 승인번호: MJU-2018-02-001-02).

3.1 피험자와 실험재료

성별과 나이에 따라서 뇌파의 차이가 있다고 알려졌기 때문에 피험자의 자격을 오른손잡이 20대의 남자로 제한하여 M대학교 학부생과 대학원생 23명을 모집하였다(Evans, Cui and Starr 1995). 실험재료는 4개의 짧은 다큐 영상물들(1분 9초~3분 5초)과 각 영상물에서 추출한 쇼트들로 구성되었다. 연구자들이 비디오 주제를 얼마나 잘 나타냈는지에 따라서 적합, 부분적합 및 부적합하다고 판정한 세 종류의 비디오 쇼트들을 동일한 비율로 선정하였다(〈표 1〉 참조).

〈표 1〉 실험에 사용된 비디오 목록

비디오 번호	표제	비디오 쇼트 수
1	검안경	9 (3: 적합, 3: 부분적합, 3: 부적합)
2	엄복동의 자전거	9 (3: 적합, 3: 부분적합, 3: 부적합)
3	백범 김구의 회중시계	15 (5: 적합, 5: 부분적합, 5: 부적합)
4	수라간 요리사	15 (5: 적합, 5: 부분적합, 5: 부적합)

3.2 실험 절차

피험자들에게 각 비디오를 시청하게 하였다. 이때 뇌파 측정과 분석을 좀 더 정확하게 수행하기 위해서 비디오의 각 쇼트가 시작하기 전에 200ms 동안 제시되는 블랙 스크린을 포함시켰다. 이후 각 비디오에서 추출한 비디오 쇼트들을 제시한 후 현재 비디오 쇼트를 단서로 하여 방금전 시청한 비디오 주제와의 적합도(비적합: 1, 부분적합: 2, 적합: 3)를 평가하도록 하였다. 이러한 평가 방법을 사용한 이유는 피험자들에게 현재 시청하고 있는 영상물 내용을 동시에 평가하게 하면 평가 결과를 입력하는 버튼 클릭과 같은 동작에 의해서 행동뇌파(motor EEG)가 발생할 수 있어 타당성을 저해하게 될 것이기 때문이다(Luck 2014).

3.3 뇌파측정과 분석

뇌파 측정을 위해서 Neuroscan Synamp amplifier(Compumedics Neuroscan, Victoria, Australia)와 Ag-AgCl electrodes가 10~20 체계에 의해 심어져 있는 32채널 Quick Cap을 사용하였다. 사용된 전극들에는 FP1/FP2, F3/Fz/F4, F7/F8 등을 포함하였고 기준(reference) 전극으로 A1과 A2를 측정하였다. 동영상 및 이미지 제시와 뇌파측정의 시작점 동기화를 통제하기 위해서 심리학 관련 실험에서 사용하는 소프트웨어인 E-Prime을 이용하였다. 이와 같이, 피험자들이 비디오를 시청하는 동안 사진에 피험자의 두피에 장착한 뇌파측정기를 통하여 뇌파를 측정하였다.

측정된 뇌파자료는 CURRY 7.09(Compumedics

Neuroscan) 프로그램을 이용하여 분석하였다. 23명의 피험자들에게 수집한 뇌파 데이터 중에서 뇌파 데이터의 품질이 낮은 1명의 뇌파 데이터를 제외시켜 22명의 데이터를 분석하였다. 앞에서 기술한 대로 뇌파분석은 피험자들이 비디오 전체를 시청할 때의 뇌파와 비디오 시청 후 제시된 비디오 쇼트를 시청할 때의 뇌파와 구분하여 분석하였다. 예를 들어서, 피험자가 특정 자극을 주제에 적합하지 않다고 판정했다고 가정해 보자. 이 경우 비디오 내의 해당 비디오 쇼트(자극)의 시작점을 찾아서 자극 제시 전 -200ms에서 자극 제시 후 1,000ms까지의 구간의 뇌파 데이터를 주제에 적합하지 않은 쇼트를 시청할 때 나타나는 반응으로 분류하여 분석하였다. 따라서 각 에폭(분석 구간)의 길이는 1.2초가 된다. 이렇게 생성된 에폭은 자극 제시 전 -200ms 값(블랙 스크린이 제시되는 동안의 뇌파값)을 이용하여 기저선 교정을 하였다. 총 에폭 1,056개(피험자당 48개 에폭들) 중에서 40개 에폭들은 잡음으로 인하여 제거되었고 주제관련 에폭 300개, 주제부분관련 에폭 404개, 주제무관 에폭 312개로 분석되었다. 연구 가설들을 검증하고 뇌파 데이터를 분류하기 위해서 주제부분관련 에폭들은 애매성 때문에 제외시키고 612개의 에폭들만 이용하였다.

비디오 쇼트를 단서로 하여 방금전 시청한 비디오 주제와의 적합도를 평가하는 것은 이 두 단계간에 불일치성이 일어날 가능성이 발생할 수 있다. 즉, 비디오를 시청하는 동안에는 특정 쇼트가 주제에 적합하지 않다고 생각했지만 비디오 시청이 끝난 후 비디오 주제가 명확해지면 해당 쇼트가 주제에 적합하다고 판단할 가능성이 있기 때문이다. 이러한 불일치성을 측정해 보기

위해서 학습기능에 의해 데이터를 분류하는 것으로 알려진 인공신경망을 이용하였다. SPSS 23의 인공신경망의 입력 데이터는 피험자들이 비디오 시청하는 동안 비디오 쇼트에 반응하는 뇌파 데이터와 피험자들이 비디오 시청 후 해당 비디오 쇼트에 부여한 적합도이다. 분류 결과, 높은 분류 정확도(82.9%)를 보여주었다. 따라서, 비디오 시청 단계의 뇌파 반응과 적합도 평가 단계의 평가 결과간에 큰 차이가 없는 것으로 나타났다.

3.4 실험 결과

3.4.1 N400/P600 효과 검증

연구가설 1과 2의 검증을 위해서 '주제적합'과 '주제비적합'을 개인내 요인(within subject factor)으로 하는 반복측정 t-검증을 실시하였다. 이러한 t-검증 결과는 주제적합 대 주제비적합 쇼트에 대한 뇌파 반응에서 N400 및 P600 효과를 나타냈으며 구체적으로 살펴보면 다음과 같다.

첫째, N400 효과에 대한 연구 가설 1의 검증 결과, 주제비적합 쇼트에 대한 N400의 음전위 최소값은 주제적합 쇼트의 음전위 최소값 보다 훨씬 크게 나타났다. 먼저 유의도 수준 0.05에서, 발현 범위는 두정엽(P8), 중앙 전두엽(Fz) 및 FC4(전두-중앙엽)로 나타났다. 유의도 수준 0.01에서, 발현 범위는 전전두엽(FP1, FP2), 전두엽(F4, F8), 중앙엽(C4) 및 전두-측두엽(FT8)에 이르기 까지 널리 걸쳐 있다. 이 중 세계 채널(FP1, F8, FT8)의 통계적 차이는 본페로니 수정(Bonferroni correction)($p < 0.0017$) 이후에도 남았다. 따라서 연구가설 1이 채택되었다.

둘째, P600 효과에 대한 연구 가설 2의 검증 결과, 주제적합 쇼트에 대한 P600의 양전위 최대값은 주제비적합 쇼트의 양전위 최대값 보다 훨씬 크게 나타났다. 유의도 수준 0.05에서, 발현 범위는 전전두엽(FP1), 전두엽(F4) 및 전두-측두엽(FT8)으로 나타났다. 유의도 수준 0.01에서는, 전두엽(F8)으로 나타났고, F8의 통계적 차이는 본페로니 수정 이후에도 남았다. 따라서 연구가설 2가 채택되었다.

3.4.2 판별 분석

앞에서 기술한 인공신경망은 EEG 신호 분류에 사용되어 비교적 높은 분류 정확도를 나타내지만 명시적인 입력 최적화 기능이 없고 학습의 결과를 쉽게 해석 할 수 없다는 단점들이 있다 (Yang et al. 2012). 따라서 ERP/MMR 모형에 기반한 비디오 스킴을 구성하기 위해서 EEG 신호 분류에서 성공적으로 사용되고 있는 SPSS 23의 판별분석을 사용하였다(Xu et al. 2014). 판별분석은 각 비디오 쇼트의 주제 적합도에 대한 판별점수를 얻을 수 있게 해 준다. 즉, 각 비디오 쇼트의 주제 적합도에 대한 평균 판별점수는 총 22명의 피험자들의 ERP 신호 분석을 통해 측정하였다. 본 연구에서는 판별함수를 구하기 위해서 단계별로 변수를 선택하는 단계별 판별분석 방법을 이용했다. 판별분석을 적용할 때 입력 데이터를 훈련 데이터와 테스트 데이터로 구분할 수 있다. 이때 훈련 데이터는 판별함수를 도출하는데 사용되고 테스트 데이터는 도출된 판별함수에 기초하여 새로운 데이터를 분류하는데 사용된다.

본 연구에서는 612개 예폭(사례)들을 교차 타당화(cross-validation) 방법에 의해서 분류

터를 기준으로 하여 나머지 8개 비디오 쇼트간의 유사도를 측정하여 가장 높은 ERP/MMR 점수(0.42)를 갖는 비디오 쇼트 S_4 가 세 번째로 선정되었다. 이러한 절차는 원하는 비디오 쇼트를 얻을 때까지 반복된다. 비디오 3의 경우는 5개의 비디오 쇼트들이 선택될 때까지 이 과정을 반복하였다(〈그림 3〉, 〈그림 4〉 참조).

내재적 및 외재적 평가를 위해서 미디어 전문가에 의해서 구성된 표준 비디오 스킴과 SBD 방법에 기반한 비디오 스킴을 구성하였다.

4.2.1 표준 비디오 스킴

세 명의 미디어 전문가들이 다음과 같은 지침에 의해서 표준 비디오 스킴들을 구성하였다 (Over, Smeaton and Awad 2008; Kim and Kim 2010).

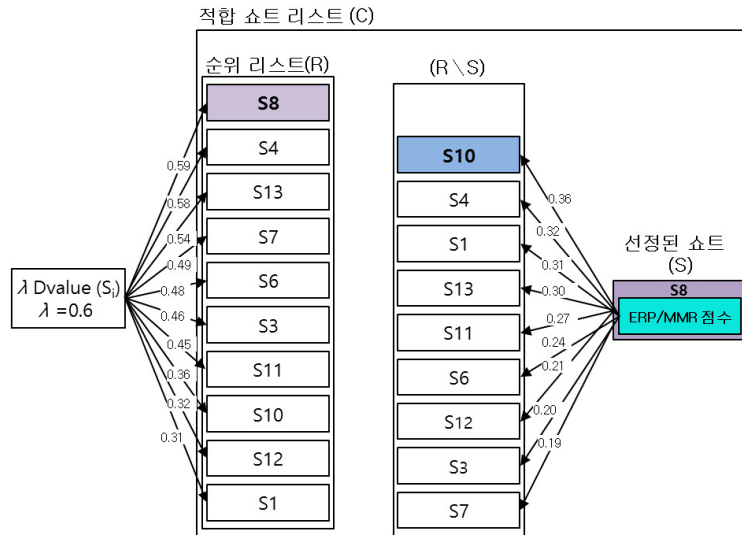
4.2 표준 비디오 스킴과 SBD 기반 비디오 스킴의 구성

ERP/MMR 모형에 기반한 비디오 스킴의

- 1) 주제와 관련된 자막이나 텍스트(예, 객체에 표시된 글자들)를 갖는 프레임들을 포

쇼트		s1	s3	s4	s6	s7	s8	s10	s11	s12	s13
프레임/단어											
f1		1	0	0	0	0	0	0	0	0	0
f2		1	0	0	0	0	0	0	0	0	0
f3		1	0	0	0	0	0	0	0	0	0
f4		0	0	0	0.87	1	1	0	0	0	0.85
f5		0	0.9	1	0	0	0.83	0	0.75	0	0.83
f6		0	0	0	0.84	0.85	0	0	0	0	1
f7		0	0	0	0	0	0	1	1	0	0
f8		0	0	0	1	0	0	0	0	0	0
f9		0	0	0	0	0	0	1	0	0	0
f10		0	0	0	0	0	0	0	0	1	0
w1	김구	1	1	1	0	0	0	0	1	1	1
w2	독립운동가	1	0	0	0	0	0	0	0	0	0
w3	김구 기념관	1	0	0	0	0	0	0	0	0	0
w4	백범일지	0	0	0	0	0	0	0	0	1	0
w5	서울	1	0	0	0	0	0	0	0	0	0
w6	용산	1	0	0	0	0	0	0	0	0	0
w7	윤봉길	0	0	0	0	0	0	1	1	1	1
w8	회중시계	0	1	1	1	1	1	0	1	1	1
w9	홍구 공원	0	0	0	0	0	0	0	0	0	1
w10	군인	0	0	0	0	0	0	1	0	0	0
Dvalue (S_i)		0.52	0.77	0.97	0.8	0.81	0.98	0.6	0.75	0.53	0.9
λ Dvalue (S_i)		0.31	0.46	0.58	0.48	0.49	0.59	0.36	0.45	0.32	0.54
$(1 - \lambda)$ Sim (S_i, S_j)		0	0.26	0.26	0.24	0.3	-	0	0.18	0.11	0.24
ERP/MMR 점수		0.31	0.2	0.32	0.24	0.19	-	0.36	0.27	0.21	0.3

〈그림 2〉 비디오 3의 ERP/MMR($\lambda=0.6$) 점수(1)



<그림 3> 비디오 3의 ERP/MMR($\lambda=0.6$) 점수(2)

함한 쇼트들을 선택한다. 만약 이 조건에 부합된 쇼트들이 많다면 비디오의 전반부 또는 후반부에 있는 쇼트들을 우선적으로 선택한다.

- 2) 주제와 관련된 객체, 사건/행동 및 인물을 갖는 프레임들을 포함한 쇼트들을 선택한다. 특히 인물의 경우에는 더 많은 수의 핵심 인물들을 포함한 프레임들을 포함하고 있는 쇼트들을 우선적으로 선택한다.

세 명의 미디어 전문가들이 독자적으로 일곱 개의 비디오에 대한 표준 비디오 스킴들을 구성하였는데 만약 세 사람 중 두 사람 이상이 한 비디오에 동일한 쇼트를 선택하면 이 쇼트는 바로 해당 비디오의 스킴으로 이용되었다. 총 일곱 개의 비디오에 대하여 25개의 쇼트들이 선정되었는데 이 중 12개 쇼트(48.0%)는 세 전문가에 의해서 선정되었고, 5개의 쇼트(20.0%)는 두 전문가에 의해서 선정되었다. 서로 의견이

일치하지 않은 나머지 8개의 쇼트들에 대해서 세 전문가들이 의논하여 최종적으로 쇼트들을 선정하여 비디오 스킴들을 구성하였다.

4.2.2 SBD 기반 비디오 스킴

SBD 방법은 일관된 높은 정확성에도 불구하고 유사한 프레임 또는 쇼트를 추출하는 단점이 있다. 따라서 이러한 문제점을 최소화하기 위해서는 컬러 히스토그램을 채택한 SBD 방법을 사용하여 비디오 쇼트들을 추출한 후 유사한 비디오 쇼트들은 그룹화하였다. 이후 각 비디오 쇼트 그룹에서 무작위로 비디오 쇼트를 추출하는 방법을 이용하였다. 예를 들어서, 비디오(예, 비디오 3)에 10개의 비디오 쇼트 그룹이 있고, 5개의 쇼트로 구성된 비디오 스킴을 생성하려면 각 두 개의 비디오 쇼트 그룹들에서 하나의 쇼트를 무작위로 추출하여 총 5개의 비디오 쇼트들로 구성된 비디오 스킴을 생성하는 방식이다(<그림 4> 참조).



〈그림 4〉 비디오 3에 대한 비디오 스킵들

4.3 네 가지 유형의 비디오 스킵의 구성

앞에서 기술한 비디오 스킵 알고리즘들의 특성을 좀 더 자세히 알아보기 위해서 네 개의 서로 다른 방식들을 비디오 3에 적용해 보았다(〈그림 4〉 참조). 〈그림 4〉의 첫 번째 예는 SBD 방식에 의해서 구성된 비디오 스킵이며, 네 번째 예는 미디어 전문가에 의해서 구성된 표준 비디오 스킵이다. 한편 세 번째 예는 유사한 비디오 쇼트가 없는 5개의 다양한 쇼트로 구성된 ERP/MMR($\lambda=0.6$) 기반 비디오 스킵을 구축하였다. 한편 이 비디오 스킵을 두 번째 예인 ERP/MMR($\lambda=1.0$) 기반 비디오 스킵(비디오 쇼트의 순위화는 판별 점수에만 기반하고

다양성을 고려하지 않음)과 비교해 보면 후자는 전자와 비교할 때 두 개의 비디오 쇼트들(S_7 and S_{13})이 동일한 두 사람들이 있는 시각적으로 비슷한 프레임들을 포함하고 있다. 시청자들의 뇌파를 분석하여 자동생성한 비디오 스킵들(ERP/MMR($\lambda=0.6$)/ERP/MMR($\lambda=1.0$))과 미디어 전문가들에 의해서 구성된 표준 비디오 스킵간의 유사성이 높아 보인다.

5. 비디오 스킵 평가

제안된 비디오 스킵의 품질과 유용성을 평가하기 위하여 내재적 평가와 외재적 평가를 수

행하였다. 내재적 평가를 위해서 제안한 알고리즘에 의해서 구성된 비디오 스킴을 표준 비디오 스킴과 비교하였다. 외재적 평가에서는 제안된 알고리즘과 SBD 방식에 의해서 구성하는 비디오 스킴들을 피조사자들에게 시청하게 한 후 이들이 비디오 주제를 얼마나 정확하게 유추해 내는지를 측정하여 제안된 알고리즘을 평가하였다. 평가를 위해서 본 연구에서 사용한 4개의 비디오와 이전 연구(김용호, 김현희 2016)에서 사용한 6개 비디오 중에서 3개 비디오(피카소, 행복에 이르는 10단계, 안경)를 합하여 총 일곱 개의 비디오들을 사용하였다. 두 연구에서 사용된 비디오의 장르, 피험자들 및 실험 절차가 유사하여 이들을 함께 통합하여 사용하는 것은 문제가 없다고 판단하였다.

5.1 내재적 평가

비디오 스킴의 품질에 대한 내재적 평가를 수행하기 위해서 ERP/MMR 방법과 SBD 방법에 의해서 구성된 비디오 스킴들을 세 명의 전문가들이 구성된 표준 비디오 스킴과 ROUGE-1

과 다양성 기준에 기초하여 비교 평가해 보았다. 먼저 비교를 위해서 짧은 요약문을 평가하는데 효과적인 것으로 입증된 ROUGE-1 측정치를 이용하였다(Lin 2004). ROUGE-1 측정치를 구하기 위해서 제안된 알고리즘에 의한 비디오 스킴과 표준 비디오 스킴의 단어/프레임 벡터들을 구성한 후 이 두 벡터간의 유사도를 코사인 유사계수를 이용하여 계산하였다. 예를 들어서, 주어진 비디오 스킴에 두 개의 쇼트들이 있는 경우 그 중 한 쇼트가 표준 비디오 스킴에 있으면 그것의 유사도는 1.0이 된다. 한편 다른 비디오 쇼트가 표준 비디오 스킴에 포함되지 않은 쇼트이면 유사도를 구하기 위해서 해당 쇼트의 단어/프레임 벡터를 표준 비디오 스킴의 단어/프레임 세트로부터 벡터와 비교한다. 해당 비디오 스킴의 최종 ROUGE-1 점수는 이 두 유사도의 평균이 된다.

제안된 방법을 평가하기 위해서 비모수적 방법인 윌콕슨 부호순위 검정을 수행하였다. 검증 결과, 유의도 수준 0.05에서 ERP/MMR($\lambda = 0.6$)/ERP/MMR($\lambda = 1.0$) 방법들의 ROUGE-1 점수들(0.79, 0.71)은 각각 SBD 방법의 ROUGE-1

〈표 2〉 ERP/MMR($\lambda = 0.6$), ERP/MMR($\lambda = 1.0$) 및 SBD 방법의 내재적 평가 결과

비디오 번호	ERP/MMR($\lambda = 0.6$)		ERP/MMR($\lambda = 1$)		SBD	
	ROUGE-1	다양성	ROUGE-1	다양성	ROUGE-1	다양성
1	0.76	0.42	0.62	0.26	0.33	0.43
2	1	0.35	0.93	0.14	0.32	0.37
3	0.77	0.39	0.80	0.19	0.42	0.51
4	0.69	0.49	0.40	0.37	0.28	0.52
5	0.76	0.39	0.77	0.39	0.19	0.34
6	0.75	0.34	0.81	0.12	0.27	0.59
7	0.77	0.35	0.65	0.15	0.32	0.43
평균	0.79	0.38	0.71	0.24	0.30	0.44

점수(0.30) 보다 통계적으로 유의미한 차이를 보이며 높게 나왔다($p=0.018$ [$\lambda=0.6$], $p=0.018$ [$\lambda=1.0$]). 또한 ERP/MMR($\lambda=0.6$) 방법의 ROUGE-1 점수(0.79)는 ERP/MMR($\lambda=1.0$) 방법의 ROUGE-1 점수(0.71) 보다 높게 나타났으나 통계적으로 유의미한 차이를 보이지는 못했다($p = 0.176$).

ROUGE-1 측정외에 비디오 스킴의 다양성을 평가 기준으로 사용하여 비디오 스킴에 있는 비디오 쇼트간의 차이를 측정했다. 예를 들어서 주어진 비디오 스킴에 두 개의 쇼트가 있는 경우 프레임/단어 벡터를 사용하여 두 쇼트의 센트로이드 벡터를 계산한다. 이후 각 벡터와 센트로이드 벡터 간의 유사도를 계산하여 첫번째 쇼트 및 두번째 쇼트와의 유사도를 측정한다. 결과적으로 측정된 이 두 쇼트들의 평균 유사도를 1에서 뺀 결과 값(비유사도)을 비디오 스킴의 다양성 점수로 사용하였다. 유의도 수준 0.05에서 ERP/MMR($\lambda=0.6$) 방법의 다양성 점수(0.38)는 ERP/MMR($\lambda=1.0$) 방법의 다양성 점수(0.24) 보다 통계적으로 유의미하게 높게 나타났다($p=0.027$). 또한 SBD 방법의 다양성 점수(0.44)가 ERP/MMR($\lambda=1.0$) 방법의 다양성 점수(0.38) 보다 높게 나타났다($p=0.028$). 이외에 ERP/MMR($\lambda=0.6$) 방법의 다양성 점수(0.38)와 SBD 방법의 다양성 점수(0.44)간에 통계적으로 유의미한 차이는 없었다($p=0.091$). 이와 같이 내재적 평가에서는 연구 가설 3은 채택되었으나, 연구 가설 4는 채택되지 못했다.

5.2 외재적 평가

외재적 평가는 피조사자들을 대상으로 하여

세 종류의 비디오 스킴들이 비디오의 의미 파악에 어떤 유용성이 있는지 파악하기 위해서 실험을 통해서 수행하였다. 다음은 실험에 투입된 피조사자와 실험 절차에 대해서 설명한다.

5.2.1 피조사자와 실험 절차

피조사자는 M대학교 학부생 39명을 활용하였다. 39명의 피조사자들을 세 개의 집단으로 고르게 나눈 다음 그룹 1(13명), 그룹 2(13명) 및 그룹 3(13명)으로 구분하였다. 일곱 개의 실험 비디오에 대한 세 개의 실험용 시스템들을 구현하였다. 구체적으로 시스템상에서 그룹 1과 그룹 2에게는 ERP/MMR($\lambda=0.6$)/ERP/MMR($\lambda=1.0$) 방식들에 의해서 구현한 비디오 스킴들을 각각 시청하게 하였다. 한편 그룹 3에게는 SBD 방식에 의해서 비디오 스킴들을 시청하게 하였다. 이와 같이 각 그룹에게 비디오 스킴을 시청하게 한 후 각 비디오에 대한 유추된 제목과 요약문을 기술하게 하였다. 이때 요약문은 최소 세 문장 이상을 기술하도록 하였고, 피조사자들이 이전에 실험 비디오들을 들은 적이 있는지 비디오별로 체크하도록 하였고, 만약 시청한 적이 있다고 표시하면 이를 분석에서 제외시켰다.

5.2.2 평가 결과

각 요약문의 점수 범위를 '0-20'로 하고 5개의 기준(0-전혀 틀림, 5-조금 맞음 10-절반 맞음, 15-절반 이상 맞음 20-정확히 맞음)을 적용하고 각 표제에는 점수 범위를 '0-10'로 하고 4개의 기준(0-전혀 틀림, 2-조금 맞음, 5-절반 맞음, 10-절반 이상 맞음)을 적용하였다. 2명의 연구자가 점수를 배정한 후 2개의 점수 평균값을 계산하여 최종 점수를 할당하였다. 평가 결

과를 분산분석으로 측정하였고, 동질적인 집단으로 나타난 세 그룹간의 차이를 분석하기 위한 사후 검증에서는 LSD 방식을 이용하였다. 측정 결과, ERP/MMR 방법이 SBD 방법 보다 더 우수한 것으로 나타났으나, ERP/MMR($\lambda=0.6$) 방법과 ERP/MMR($\lambda=1.0$) 방법간에 유의미한 차이는 없는 것으로 나타났다(〈표 3〉 참조). 이와 같이 외재적 평가에서도 연구 가설 3은 채택되었으나, 연구 가설 4는 채택되지 못했다.

5.3 논의

제안된 ERP/MMR 기반 비디오 요약 알고리즘은 ERP 기반 주제 적합성 모형을 사용하여 구성되었다. 주제 적합성 모형에서 시청자의 ERP 반응은 주제와 적합하지 않은 비디오 쇼트에 대하여 N400 효과를 나타낼 것으로 가정하였다. 예상한 대로, 주제비적합 쇼트에 대한 N400의 음전위 최저값이 주제적합 쇼트의 음전위 최저값 보다 크게 나타났고, 발현 범위는 전전두엽, 전두엽 및 전두-측두엽에 걸쳐 있었다. 이러한 결과는 비디오를 실험 재료로 사

용한 김용호와 김현희(2016) 연구 및 사운드 없는 비디오를 사용한 Sitnikova et al. (2008) 연구의 결과들과 유사하다.

주제 적합성 모형은 시청자의 ERP 반응은 주제와 적합한 비디오 쇼트에 대하여 P600 효과를 나타낼 것으로 가정하였다. 주제적합 비디오 쇼트에 대한 시청자의 ERP 반응은 전두엽에서 P600 효과를 나타냈다. 이러한 결과는 Nakano 외(2014)에 의해 주장된 두뇌 전측(anterior)에 나타나는 P600은 개념적으로 복잡한 메모리 동작을 포함한 프로세스와 관련되어 있다는 주장과 일맥상통한다.

이후 ERP/MMR 기반 알고리즘의 성능을 내재적 및 외재적 평가를 통해서 측정해 보았다. 내재적 평가에서는 유의도 수준 0.05에서 ERP/MMR($\lambda=0.6$)/ERP/MMR($\lambda=1.0$) 방법들의 ROUGE-1 점수들은 각각 SBD 방법의 ROUGE-1 점수 보다 통계적으로 유의미한 차이를 보이며 높게 나왔다. 또한 ERP/MMR($\lambda=0.6$) 방법의 ROUGE-1 점수는 ERP/MMR($\lambda=1.0$) 방법의 ROUGE-1 점수 보다 높게 나타났다나 통계적으로 유의미한 차이를 보이지는 못했다. 비디오 스킴의 다양성을 측정할 결

〈표 3〉 ERP/MMR($\lambda=0.6$), ERP/MMR($\lambda=1.0$) 및 SBD 방법의 외재적 평가 결과

비디오 스킴 종류	평균	표준 편차	F	p
ERP/MMR ($\lambda=0.6$)	25.91	5.26	47.13	0.000
ERP/MMR ($\lambda=1.0$)	23.64	4.86		
SBD	15.33	6.05		
사후 검증 결과 (LSD)				
대응			유의확률	
ERP/MMR ($\lambda=0.6$) vs. SBD			0.000	
ERP/MMR ($\lambda=1.0$) vs. SBD			0.000	
ERP/MMR ($\lambda=0.6$) vs. ERP/MMR ($\lambda=1.0$)			0.057	

과, 유의도 수준 0.05에서 ERP/MMR($\lambda=0.6$) 방법의 다양성 점수는 ERP/MMR($\lambda=1.0$) 방법의 다양성 점수 보다 통계적으로 유의미한 차이를 보이며 높게 나타났다. 이외에 SBD 방법의 다양성 점수도 ERP/MMR($\lambda=1.0$) 방법의 다양성 점수 보다 통계적으로 유의미하게 높게 나타났다.

외재적 평가는 앞에서 설명한 내재적 평가와 유사한 결과를 보여 주었다. 유의도 수준 0.05에서 ERP/MMR($\lambda=0.6$) 방법과 ERP/MMR($\lambda=1.0$) 방법의 평가 점수들이 SBD 방법의 평가 점수 보다 각각 유의미한 차이를 보이며 높게 나타났다. 그러나 ERP/MMR($\lambda=0.6$) 방법과 ERP/MMR($\lambda=1.0$) 방법간에는 유의미한 차이는 없는 것으로 나타났다. 이와 같이 MMR 모형을 사용하여 비디오 스킵의 다양성을 높이는 것이 비디오의 주제 파악을 하는 데 실질적인 도움을 주지 못했다는 결과가 나왔다. 그러나 이러한 평가 결과를 확정하기 위해서는 앞으로 더 많은 검증 과정들이 필요하다고 생각한다. 본 연구의 실험에서 사용된 한 주제에 초점을 맞춘 다큐 비디오 대신 복수 개의 주제를 다룬 다큐 비디오에서는 비디오 스킵의 다양성이 주제 파악에 영향을 미치는 중요한 요인으로 작용할 수 있을 것으로 예측해 본다. 이에 대한 후속 연구가 필요해 보인다.

6. 결 론

본 연구는 비디오 스킵을 자동으로 생성하기 위한 비디오 요약 알고리즘을 제안하고 평가해 보았다. 제안된 알고리즘은 ERP-기반 주제 적

합성 모형, MMR 기법 및 판별분석기법을 사용하여 구현하였다. 이후 구현된 ERP/MMR 기반 알고리즘을 이용하여 구성된 비디오 스킵들의 품질과 유용성을 내재적 및 외재적 평가를 통해서 검증해 보았다. 내재적 평가에서는 ERP/MMR($\lambda=0.6$)/ERP/MMR($\lambda=1.0$) 방법들의 ROUGE-1 점수들은 각각 SBD 방법의 ROUGE-1 점수 보다 유의미한 차이를 보이며 높게 나왔다. 또한 비디오 스킵의 다양성을 측정된 결과, ERP/MMR($\lambda=0.6$) 방법의 다양성 점수는 ERP/MMR($\lambda=1.0$) 방법의 다양성 점수 보다 유의미한 차이를 보이며 높게 나타났다. 외재적 평가 결과에서도 ERP/MMR($\lambda=0.6$) 방법과 ERP/MMR($\lambda=1.0$) 방법의 평가 점수들이 SBD 방법의 평가 점수 보다 각각 유의미한 차이를 보이며 높게 나타났다. 그러나 내재적 및 외재적 평가 모두에서 ERP/MMR($\lambda=0.6$) 방법의 평가 점수와 ERP/MMR($\lambda=1.0$) 방법의 평가 점수 간에 통계적으로 유의미한 차이는 없는 것으로 나타났다.

BCI(brain computer interface) 연구는 주목 받는 특정 자극에 대한 뇌파 반응이 다른 자극에 대한 뇌파 반응과 구분된다는 것을 증명하였다(Seoane, Gabler and Blankertz 2015). 따라서 비디오 스킵을 생성하기 위해 본 연구에서 제안된 방법은 EEG 장치를 장착한 시청자들에게 비디오들을 보여준 후 그들의 ERP 신호를 기반으로 비디오 주제에 적합한 비디오 쇼트들을 추출할 수 있다. BCI 시스템은 현재 시청자가 보고 있는 비디오 쇼트가 주제에 적합한지 아닌지를 판단하는 분류 모듈(즉, 판별 분석의 사용)을 필요로 한다. 예를 들어서, 시스템은 특정 비디오 쇼트의 ERP 데이터가 전

두업 영역에서 P600 효과를 나타내는 경우 해당 비디오 쇼트를 추출하여 본 연구에서 제안한 ERP/MMR 모형에 기반한 비디오 스킴을 생성할 수 있다. 이렇게 생성된 비디오 스킴들의 패턴을 체계적으로 분석하여 어떤 유형의 쇼트들이 선정되었는지에 대한 연구도 필요해 보인다. 궁극적으로 이러한 비디오 스킴들과 비디오들간의 관계를 딥러닝 방식으로 분석하여 비디오 스킴을 자동으로 생성할 수 있는 인

프리를 구축할 수 있을 것으로 기대해 본다. 본 연구에서 제안한 알고리즘 즉 시청자들의 뇌파를 분석하여 자동 생성한 비디오 스킴들을 딥러닝의 훈련 데이터로 사용한 경우와 비디오 스킴 전문가(제작자)들이 수작업으로 작성한 비디오 스킴들을 훈련 데이터로 사용한 경우(예, The Morgan)와는 테스트 데이터의 결과에서 어떤 차이가 나타날 것인지에 대한 후속 연구도 흥미로워 보인다.

참 고 문 헌

- [1] 권준수. 2000. 인지 기능 연구에서의 사건관련전위의 이용. 『인지과학작업』, 1(1): 79-98.
- [2] 김용호, 김현희. 2016. 실시간 뇌파반응을 이용한 주제관련 영상물 쇼트 자동추출기법 개발연구: ERP의 N400 및 P600 가설검증을 중심으로. 『멀티미디어학회논문지』, 19(8): 1260-1274.
- [3] 정영미. 2012. 『정보검색연구』. 서울: 연세대학교 대학출판문화원.
- [4] Allegretti, M. et al. 2015. "When Relevance Judgement is Happening?: An EEG-based study." In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 719-722). New York: ACM.
- [5] Burmester, J., Spalek, K. and Wartenburger, I. 2014. "Context Updating during Sentence Comprehension: The Effect of Aboutness Topic." *Brain and Language*, 137: 62-76.
- [6] Carbonell, J. and Goldstein, J. 1998. "The Use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries." In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 335-336). New York: ACM.
- [7] Chen, F., Delannay, D. and De Vleeschouwer, C. 2011. "An Autonomous Framework to Produce and Distribute Personalized Team-sport Video Summaries: A Basketball Case Study." *IEEE Transactions on Multimedia*, 13(6): 1381-1394.
- [8] Eugster, M. J. et al. 2016. "Natural Brain-information Interfaces: Recommending Information by Relevance Inferred from Human Brain Signals." *Scientific Reports*, 6(38580), 1-10.
- [9] Evans, W. J., Cui, L. and Starr, A. 1995. "Olfactory Event-related Potentials in Normal Human

- Subjects: Effects of Age and Gender.” *Electroencephalography and Clinical Neurophysiology*, 95(4): 293-301.
- [10] Hu, W. et al. 2011. “A Survey on Visual Content-based Video Indexing and Retrieval.” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(6): 797-819.
- [11] IBM Research Corp. 2016. *Morgan / IBM Creates First Movie Trailer by AI*. [online] [cited 2018. 8. 3.] <<https://www.youtube.com/watch?v=gJEzuYynaiw>>
- [12] Kim, H. H. and Kim, Y. H. 2010. “Toward a Conceptual Framework of Key-frame Extraction and Storyboard Display for Video Summarization.” *Journal of the Association for Information Science and Technology*, 61(5): 927-939.
- [13] Lin, C. 2004. “ROUGE: A Package for Automatic Evaluation of Summaries.” In *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*, Barcelona, Spain.
- [14] Luck, S. J. 2014. *An Introduction to the Event-related Potential Technique*. Cambridge: MIT press.
- [15] Mayer, R. E. 2009. *Multimedia Learning*. New York: Cambridge University Press.
- [16] Mehmood, I. et al. 2016. “Divide-and-conquer based Summarization Framework for Extracting Affective Video Content.” *Neurocomputing*, 174: 393-403.
- [17] Mishra, R. et al. 2015. “Real time and Non Real time Video Shot Boundary Detection using Dual Tree Complex Wavelet Transform.” *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, pp. 1495-1500). New York: IEEE.
- [18] Mostafa, J. and Gwizdka, J. 2016. “Deepening the Role of the User: Neuro-Physiological Evidence as a Basis for Studying and Improving Search.” In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval* (pp. 63-70). New York: ACM.
- [19] Mudrik, L. et al. 2014. “Synchronous Contextual Irregularities Affect Early Scene Processing: Replication and Extension.” *Neuropsychologia*, 56: 447-458.
- [20] Nakano, H. et al. 2014. “Electrophysiological Response to Omitted Stimulus in Sentence Processing.” *NeuroReport*, 25(14): 1169-1174.
- [21] Over, P., Smeaton, A. F. and Awad, G. 2008. “The TRECVID 2008 BBC Rushes Summarization Evaluation.” In *Proceedings of the 2nd ACM TRECVID Video Summarization Workshop* (pp. 1-20). New York: ACM.
- [22] Schumacher, P. B. and Hung, Y. C. 2012. “Positional Influences on Information Packaging: Insights from Topological Fields in German.” *Journal of Memory and Language*, 67(2):

- 295-310.
- [23] Seoane, L. F., Gabler, S. and Blankertz, B. 2015. "Images from the Mind: BCI Image Evolution based on Rapid Serial Visual Presentation of Polygon Primitives." *Brain-Computer Interfaces*, 2(1): 40-56.
- [24] Sitnikova, T. et al. 2008. "Two Neurocognitive Mechanisms of Semantic Integration during the Comprehension of Visual Real-world Events." *Journal of Cognitive Neuroscience*, 20(11): 2037-2057.
- [25] Tavassolipour, M., Karimian, M. and Kasaei, S. 2014. "Event Detection and Summarization in Soccer Videos using Bayesian Network and Copula." *IEEE Transactions on Circuits and Systems for Video technology*, 24(2): 291-304.
- [26] Wang, L. and Schumacher, P. B. 2013. "New is not always Costly: Evidence from Online Processing of Topic and Contrast in Japanese." *Frontiers in Psychology*, 4: 363.
- [27] Xu, R. et al. 2014. "Enhanced low-latency Detection of Motor Intention from EEG for Closed-loop Brain-computer Interface Applications." *IEEE Transactions on Biomedical Engineering*, 61(2): 288-296.
- [28] Yang, J. et al. 2012. "Channel Selection and Classification of Electroencephalogram Signals: An Artificial Neural Network and Genetic Algorithm-based Approach." *Artificial Intelligence in Medicine*, 55(2): 117-126.
- [29] Zhu, X. et al. 2007. "A Text-to-picture Synthesis System for Augmenting Communication." In *Proceedings of the 22nd National Conference on Artificial Intelligence-Volume 2* (pp. 1590-1595). Menlo Park, CA: AAAI Press.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Kwon, Jun Soo. 2000. "The Use of Event-related Potentials in the Study of Cognitive Functions." *Journal of Cognitive Science*, 1(1): 79-98.
- [2] Kim, Hyun-Hee and Kim, Yong-Ho. 2016. "Automatic Extraction Techniques of Topic-Relevant Visual Shots using Real Time Brainwave Responses: ERP N400 and P600 Hypotheses Test." *Journal of Korea Multimedia Society*, 19(8): 1260-1274.
- [3] Chung, Young-Mee. 2012. *Research in Information Retrieval*. Seoul: Yonsei University Press.