

뇌파측정기술을 활용한 언어 기반 사운드 요약의 생성 방안 연구*

Towards the Generation of Language-based Sound Summaries Using Electroencephalogram Measurements

김현희 (Hyun-Hee Kim)**

김용호 (Yong-Ho Kim)***

초 록

본 연구는 시청자가 사운드 자료의 주제를 파악하는 과정과 사운드의 특성을 이해하기 위한 인지적 정보처리 모형을 구성하였다. 이후 사건관련유발전위(event related potentials, ERP)의 두뇌의 전후측에 걸쳐서 발현하는 N400, P600 구성요소들을 인지적 정보처리 모형의 언어적 표상에 접목시켜 사운드 요약을 생성하는 방안을 제안하기 위해서 연구 가설들을 수립하였다. 뇌파 실험을 통해서 연구 가설들을 검증한 결과, P600이 사운드 요약의 핵심 구성 요소로 나타났다. 본 연구 결과는 분류 알고리즘 설계에 적용되어 내용 기반 메타데이터 즉, 일반적인 또는 개인화된 미디어 요약(사운드 요약, 비디오 스킴)을 생성하는 데에 활용될 수 있을 것이다.

ABSTRACT

This study constructed a cognitive model of information processing to understand the topic of a sound material and its characteristics. It then proposed methods to generate sound summaries, by incorporating anterior-posterior N400/P600 components of event-related potential (ERP) response, into the language representation of the cognitive model of information processing. For this end, research hypotheses were established and verified them through ERP experiments, finding that P600 is crucial in screening topic-relevant shots from topic-irrelevant shots. The results of this study can be applied to the design of classification algorithm, which can then be used to generate the content-based metadata, such as generic or personalized sound summaries and video skims.

키워드: 인지적 모형, 사건관련유발전위, 전후측 N400, 전후측 P600, 사운드 요약, 비디오 스킴, 인공신경망, 내용 기반 메타데이터
cognitive model, event-related potential response, anterior-posterior N400, anterior-posterior P600, sound summaries, video skims, artificial neural networks, content-based metadata

* 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: 2019년 8월 19일 ■ 최초심사일자: 2019년 9월 17일 ■ 게재확정일자: 2019년 9월 24일
■ 정보관리학회지, 36(3), 131-148, 2019. <http://dx.doi.org/10.3743/KOSIM.2019.36.3.131>

1. 서 론

1.1 연구 배경과 목적

최근 주제나 내용을 전달하는 핵심 채널로 언어를 사용하고 동영상상을 보조적 채널로 사용하는 비디오 자료들이 소셜미디어, 디지털 도서관 등의 핵심 자원으로 빠르게 증가해 가고 있다. 특히 유튜브에서 이러한 유형의 교육 자료의 증가는 2017년에 70%에 이르렀고(Martin, 2018), 많은 연구들(Alwehaibi, 2015; Buzzetto-More, 2015)은 이러한 교육 자료가 높은 학습 효과를 나타냈다고 보고하고 있다.

언어 중심의 비디오가 교육 콘텐츠로 중요해지고 있음에도 불구하고 디지털도서관이나 소셜미디어 플랫폼들은 정보검색에 필요한 충분한 메타데이터 없이 자료를 제공하고 있다. 이로 인하여 이용자들은 이들 미디어 자료의 전체 또는 원하는 부분만을 활용하는 것이 어렵다. 특히 이용자가 이러한 미디어를 실제 이용할 지 여부를 결정할 때 참조한다고 알려진 사운드 요약, 비디오 스킵과 같은 내용 기반 메타데이터가 충분히 제공되지 못하고 있다.

본 연구는 언어 중심의 비디오에서 주제 또는 내용 파악에 도움을 주는 스피치 부분을 추출하여 사운드 요약을 생성하는 방안을 제안하고자 한다. 사운드 요약에 대한 기존 연구들을 살펴보면, 잠재의미분석 방법(단어간의 관계, 단어와 문장간의 관계, 문장간의 관계를 분석하여 핵심 문장들을 추출)을 활용한 연구가 있다(Kim & Kim, 2016). 이외에 사운드의 음향학적 자질 즉, 사운드의 속도, 피치(사운드의

높낮이), 강도(사운드의 세기)를 사운드 요약에 적용하는 연구들도 있다(Maskey & Hirschberg, 2006; Zhang & Fung, 2012).

이러한 기존 방법들은 사운드의 외적 요인 중 하나인 이용자(시청자)들이 사운드의 내용(또는 주제)를 이해하는 과정을 고려하지 않고 사운드의 내적 특성에만 초점을 맞추었다. 따라서 이용자가 사운드 요약을 보고 생각하는 사운드의 주제와 실제 사운드의 주제간에 차이가 생겨날 수 있다. 예를 들어서, 사운드의 음향학적 자질을 사용하는 경우에는 이용자의 주목을 끌 수 있는 사운드 자질을 가려낼 수 있지만, 주목을 끈 자질이 주제적으로 중요한 자질인지에 대한 문제는 여전히 남으며, 이러한 문제점은 의미격차 문제와 밀접하게 관련된다(Hu, Xie, Li, Zeng, & Maybank, 2011). 따라서 이러한 의미 격차를 완화시키기 위해서 시청자가 어떤 인지적 과정을 거쳐 주제적으로 중요한 사운드 내용을 파악하는지에 대한 이해가 선행되어야 한다.

이에 따라 본 연구의 목적은 다음과 같다. 첫째, 스피치와 동영상으로 구성된 비디오에서 각 채널의 특성 및 이 두 채널간의 상호 작용 효과를 이해하기 위한 인지적 정보처리 모형을 구현한다. 둘째, 구현된 인지적 정보처리 모형의 언어적 표상과 언어와 영상이 결합된 통합적 표상에 각각 뇌파측정기술을 접목시켜 자료의 주제나 핵심 내용을 담고 있는 스피치 부분을 추출하는 방안을 제안한다. 셋째, 제안된 방안을 일반적인 또는 개인화된 미디어의 내용 기반 메타데이터 생성의 기초 기술로 활용하는 데에 연구의 목적을 두었다.

1.2 용어 정의

(1) 사운드 요약: 유튜브에서 자막화면에 사운드를 입힌 형태의 비디오들이 널리 이용된다. 이런 형태의 비디오를 요약하는 방식으로 청각 채널의 말과 시각채널의 자막만을 이용하는 비디오 요약형태이다. 본 연구에서는 언어 중심의 비디오에서 동영상을 제거하고 자료의 주제나 핵심 내용을 담고 있는 스피치 부분을 추출한 후 추출된 스피치 내용을 자막화하여 스피치와 자막을 함께 사용한다.

(2) 사건관련유발전위(event related potentials, ERP): 인간의 두뇌 외피에 전극을 부착하여 외부의 자극 또는 내부의 심리적 과정과 관련되어 나타나는 뇌의 반응으로 일정한 시간 동안의 전기적 활동을 의미한다. ERP의 구성요소로 N400, P600 등이 있다.

(3) 기저선(baseline): 뇌파 실험에서 독립변인의 처치 효과를 측정하기 위해 처치를 가한 상태의 조건과 비교하기 위해 사용하는 일상적 행동을 지칭한다. 본 연구에서는 쇼트(자극) 제시 전 이전 쇼트의 마지막 200ms 동안의 평균 진폭을 기저선으로 정하였다.

(4) 에폭(epoch): 뇌파 실험에서 분석할 뇌파의 에폭 구간으로 본 연구에서는 자극 제시 전 200ms부터 자극 제시 후 1,000ms로 설정하여 각 에폭의 길이는 총 1.2초로 정하였다.

(5) 전후측(anterior/posterior) N400: 전측 N400은 자극 제시 이후 400ms 근처에서 전두엽, 중앙엽에서 음전위(negative) 최저값으로 나타나며, 후측 N400은 자극 제시 이후 400ms 근처에서 중앙엽, 두정엽에서 음전위 최저값으로 나타난다.

(6) 전후측 P600: 전측 P600은 자극 제시 이후 600ms 근처에서 전두엽, 중앙엽에서 양전위(positive) 최고값으로 나타나며, 후측 P600은 자극 제시 이후 600ms 근처에서 중앙엽, 두정엽에서 양전위 최고값으로 나타난다.

(7) 일반적인 개인화된(generic, personalized) 미디어 요약: 일반적인 미디어 요약은 다수의 이용자들을 대상으로 구성한 각 미디어에 할당된 하나의 요약이며, 개인화된 미디어 요약은 특정 개인의 취향에 따라 구성된 맞춤형 미디어 요약이다.

1.3 연구 방법

첫째, 인지적 정보처리 모형을 구성하기 위해서 Mayer(2005)의 모형과 Baddeley(2007)의 모형을 살펴보고, ERP에 관한 선행연구들도 분석하였다.

둘째, 인지적 정보처리 모형을 뇌파측정기술에 접목시켜 사운드 요약을 구성하는 방안을 구현하기 위해서 연구 가설들을 수립하고 이 연구 가설들을 뇌파 실험을 통해서 검증하였다. 뇌파 실험은 언어 중심의 비디오내의 사운드 효과를 살펴보기 위해서 수행되었다.

셋째, 피험자들은 M대학, Y대학의 학부생(34명)과 대학원생(2명)의 36명으로 무작위로 실험집단과 통제집단으로 나누었다. 실험집단(18명)에게는 사운드 실험자료(스피치+자막)를 사용하였고, 통제집단(18명)에게는 비디오 실험자료(스피치+동영상)를 사용하였다.

넷째, 뇌파 측정을 위해서 Neuroscan Synamp amplifier(Compumedics Neuroscan, Victoria, Australia)의 30채널 Quick Cap과 E-Prime을

이용하였다. 측정된 뇌파자료는 CURRY 7.012 (Compumedics Neuroscan) 프로그램을 이용하여 분석되었다. SPSS 23의 반복측정(repeated measures) MANOVA, 반복측정 t-검증, 판별 분석 및 인공신경망(artificial neural networks) 분석을 이용하여 분석 결과를 통계적으로 검증하였다.

2. 이론적 배경

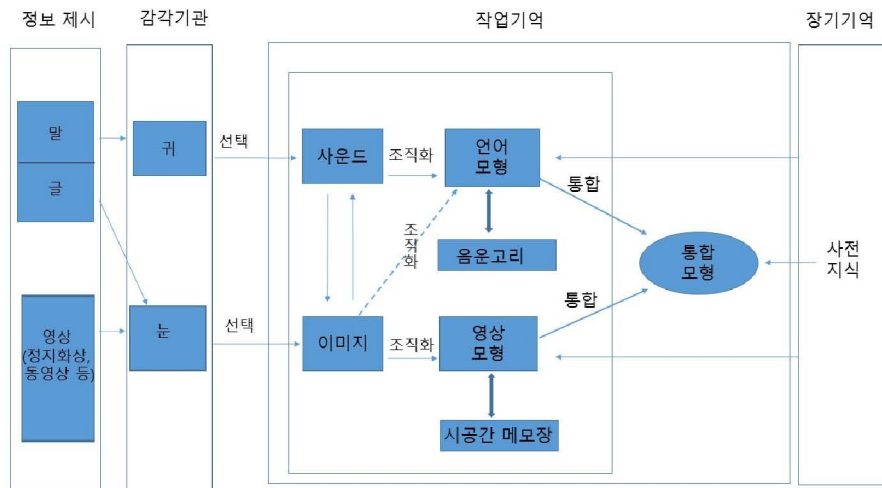
2.1 인지적 정보처리 모형

Mayer(2005)은 정보를 처리할 때 두 개의 채널인 언어 정보와 영상 정보가 어떻게 처리되는지에 대한 자세한 설명을 그의 멀티미디어 학습 모형에서 제시하지 않았다. 따라서 이러한 점을 보강하기 위해서 본 연구에서 제안한 인지적 정보처리 모형은 Mayer 모형과 Baddeley 모형을 기반으로 하여 구성되었다. 본 연구에서 제안한 모형의 특성은 다음과 같다. 첫째, Mayer 모형에서 주장한 것처럼 사람들은 정보를 처리할 때 두 개의 채널 즉, 시각 정보와 청각 정보를 처리하는 절차가 다르게 코딩되어 있다고 가정하는 이중채널, 각 채널의 제한된 용량, 능동적 처리의 세 가지 기본적인 사항을 수용하였다. 다만 차이점은 자극이 '글'인 경우 사운드로 바꾸지 않고 바로 언어적 표상으로 구조화되는 경우도 가능하기 때문에 이를 모형에 추가하였다(〈그림 1〉의 점선 참조). 둘째, Mayer 모형의 작업기억에서 언어 모형과 영상 모형의 동적인 정보 처리 과정을 기술하기 위해서 Baddeley 모형의 음운고리(phonological loop)(사운드 정

보를 저장하고 저장된 정보를 다시 사운드화하여 정보를 재활성화시킴)와 시공간 메모장(visuospatial sketchpad)(시각계에서 받아들이는 시공간 정보나 언어정보에서 만들어진 시각적 이미지 정보를 저장하고 조작함)을 이 두 모형들에 각각 연계시켰다(Hakoda, 2010)(〈그림 1〉의 두 개의 굵은 실선 참조).

인지적 정보처리 모형에서 글(printed text)과 말(spoken text)이 언어 형태로 제시된다(〈그림 1〉 참조). 말은 감각기관으로 가서 분석되지 않은 사운드의 형태로 표상되며 청각적 감각기관을 통과할 때 메시지에 포함되어 있는 관련된 중요한 몇 개의 문자(예, 단어)에 주의를 집중한다. 다시 말해서 청각적 처리를 담당하는 채널의 제한된 용량으로 인하여 모든 문자들을 수용하지 못하고 일부 문자를 선택하게 된다. 이 때 사운드 이미지에서 하나의 시각적 이미지로 전환될 수 있다(〈그림 1〉의 화살표 중 작업기억의 좌측에 있는 '사운드'에서 '이미지'로 향하는 부분을 말함). 즉 '강아지'라는 사운드를 들었을 때 이 사운드에 의해 연상되는 시각적 표상인 '강아지' 이미지를 의미한다. 이후 작업기억에서 선택된 문자들은 음운고리 장치를 거쳐서 언어 모형이라는 하나의 종합된 표상으로 구조화된다.

글은 시청각 감각기관을 통과하여 작업기억에서 시각적 이미지에서 하나의 사운드로 전환될 수 있다(〈그림 1〉의 화살표 중 작업기억의 좌측에 있는 '이미지'에서 '사운드'로 향하는 부분을 말함). 즉, 하나의 시각적 이미지('강아지'라는 문자)가 '강아지'라는 문자에 대한 사운드로 전환되어 음운고리를 거쳐서 언어적 표상으로 구조화되거나 사운드로의 전환을 거치지 않고 바로 언어적 표상으로 구조화될 수 있다.



〈그림 1〉 인지적 정보처리 모형(Baddeley, 2007; Mayer, 2005)

한편 영상이란 삽화, 도표, 사진과 같은 정지 화상, 동영상을 나타낸다. 영상은 감각기관으로 가서 분석되지 않은 이미지의 형태로 표상되며 시각적 감각기관을 통과할 때 시각적 처리를 담당하는 채널의 제한된 용량으로 인하여 관련된 중요한 이미지 또는 이미지의 중요한 부분을 선택하게 된다. 이후 작업기억에서 선택된 이미지 또는 이미지 조각은 시공간 메모장 장치를 거쳐서 영상 모형이라는 하나의 종합된 표상으로 구조화된다. 만약 제시된 정보(자극)가 비디오(스피치/동영상)인 경우 언어적 표상과 영상적 표상이 서로 통합된 표상(통합 모형)과 장기기억에서 이미 가지고 있는 사전 지식과 서로 연계되어 능동적으로 처리된다.

이와 같이 통합 처리되는 과정을 비디오(스피치+동영상)를 이용하여 설명하면 다음과 같다. 인지적 정보처리과정의 초기 지각과정에서 주목기능이 중요하고, 통합과정에서는 추론기능이 중요하다고 할 수 있다. 각 채널의 제한된 용량으로 인하여 인지체계는 감각기관에 입력

되는 모든 감각정보를 수용하지 못하며 중요한 감각정보를 선별하여 작업기억에 투입시킨다고 가정할 수 있다. 비디오 내용에 대한 의미추론과정은 시각채널을 통해 유입되는 시각정보와 청각채널을 통해 유입되는 청각정보의 통합 과정이며 이는 작업기억 내에서 이루어진다. 이 통합과정에서 새로운 감각정보를 의미적 추론의 대상으로 삼기 위해서는 장기기억에 저장된 지식의 도움이 필요하다. 장기기억 속의 지식은 호출되어 감각정보 내용에 대한 사실 검증은 물론 비디오 내용의 추론이라는 논리적 작업의 일관성을 관리하는 기준이 된다. 또한 현재 시청하고 있는 비디오 쇼트를 의미적으로 처리하기 위해서는 해당 쇼트 이전까지의 쇼트들을 시청하여 파악한 정보에 대한 기억도 포함된다.

2.2 선행 연구와 연구 가설 설정

사운드 실험자료의 경우 말(스피치)과 글(자

막)이 피험자에게 제시되면 피험자는 주제적으로 중요한 말과 글에 집중한다. 이후 작업기억에서 선택된 말과 글은 조직화를 통해서 하나의 언어적 표상으로 구조화되고 이러한 언어적 표상이 장기기억에서 이미 가지고 있는 사전 지식과 통합되어 능동적으로 처리된다고 가정한다(<그림 1>의 언어 모형 참조).

비디오 실험자료의 경우 말과 동영상이 피험자에게 제시되면 피험자는 주제적으로 중요한 말과 동영상에 집중한다. 이후 말은 앞에서 기술한 것처럼 작업기억에서 선택과 구조화를 통해서 언어적 표상이 되고, 동영상은 선택과 구조화를 통해서 영상적 표상으로 구조화된다. 이후 언어적 표상과 영상적 표상이 통합된 후 이 통합적 표상이 장기기억에서 이미 가지고 있는 사전 지식과 연계되어 처리된다고 가정한다(<그림 1>의 통합 모형 참조).

2.2.1 연구 가설 1: 전후측 N400

스피치와 동영상이 사용된 비디오 실험자료의 경우 해당 쇼트가 주제에 적합하지 않을 경우 N400 효과를 나타낼 것이며, 발현 범위는 전두엽, 중앙엽 영역으로 두뇌의 전측이 될 것으로 가정한다. 또한 스피치와 자막이 사용된 사운드 실험자료의 경우 해당 쇼트가 주제에 적합하지 않을 경우 N400 효과를 나타낼 것이다. 발현 범위는 두뇌의 중앙엽, 두정엽 영역으로 두뇌의 후측이 될 것으로 가정한다. 이외에 사운드의 경우 스피치와 자막이 함께 처리되어 서로 보완 효과가 생겨나 언어 처리와 관계되는 것으로 알려진 측두엽 영역에서 강한 N400 효과가 나타날 것으로 가정한다.

이러한 주장의 근거는 다음과 같다. 많은 연구

들(Nieuwland & Martin, 2012; van Berkum, Hagoort, & Brown, 1999)은 의미적으로 또는 문법적으로 맞지 않는 단어(또는 문장)의 자극에 대해서 후측 N400 효과를 보인다고 보고하였다. Wilson, Bautista, McCarron(2018)은 말로 표현된 (또는 인쇄된) 문자 처리는 상측 측두엽(superior temporal sulcus) 영역에서 수행된다고 주장하였다. 비디오에 적합한 연구들을 살펴보면, Kutas와 Federmeier(2011)은 영화 장면의 자극이 단어의 자극보다는 상대적으로 두뇌의 앞쪽에서 발현되는 전측 N400 효과를 나타내는 경향이 있다고 보고하고 있다. Kim과 Kim(2019a; 2019b)은 동영상과 스피치가 동일한 비중으로 주제를 표현하는 데에 사용된 다큐멘터리 비디오들을 실험 자료로 사용했을 때 비디오 주제에 적합하지 않는 자극(비디오 쇼트)에 대해서 전측 N400 효과가 나타난다고 주장했다. 본 연구에서 실험 데이터로 사용하는 비디오는 스피치에 의해서 주로 정보를 제공하기 때문에 측두엽 영역도 활성화될 것이라고 예측할 수 있다. 그러나 스피치와 함께 제공되는 동영상으로 인하여 그 효과가 상쇄될 수 있다고 가정하여 비디오에서는 측두엽 영역을 활성화 범위에서 제외시켰다. 이에 따라서 연구 가설 1은 다음과 같이 두 개로 세분화된다.

- (1) 연구 가설 1-1: 비디오 실험자료의 주제 비적합 쇼트는 전측에서 N400 효과를 나타낼 것이다.
- (2) 연구 가설 1-2: 사운드 실험자료의 주제 비적합 쇼트는 후측 및 측두엽에서 N400 효과를 나타낼 것이다.

2.2.2 연구 가설 2: 전후측 P600

비디오 실험 자료의 경우 해당 쇼트가 비디오 주제에 적합할 경우 P600 효과를 나타낼 것이며, 발현 범위는 두뇌의 전측이 될 것으로 가정한다. 또한 사운드 실험 자료의 경우 해당 쇼트가 주제와 관련이 있는 경우 P600 효과를 나타낼 것이며, 발현 범위가 두뇌의 후측이 될 것이다. 즉, 해당 쇼트가 주제에 관련된 경우 이를 기존 잠정적인 주제와 비교하여 현재의 주제를 그대로 유지할 지 또는 새로운 잠정적인 주제를 설정할 지와 같은 문맥 갱신 과정을 거치게 되며 이때 P600 효과가 나타난다는 것이다. 이외에 사운드 실험 자료의 경우 측두엽 영역도 활성화 될 것으로 가정한다.

이러한 주장의 근거를 살펴보면 다음과 같다. P600은 일반적으로 문맥-갱신 단계(Burmester, Spalek, & Wartenburger, 2014), 담화의 내적인 재조직과 통합(Wang & Schumacher, 2013)과 관련되어 있다고 보고되고 있다. 많은 연구들(DeLong, Quante, & Kutas, 2014; Geyer, Holcomb, Kuperberg, & Perlmutter, 2006)은 언어를 자극으로 사용할 경우 후측 P600 효과가 나타난다고 주장하였다. 한편 Nakano, Rosario, Oshima-Takane, Pierce와 Tate(2014)는 전측 P600은 개념적으로 복잡한 기억 작동을 포함한 과정과 관련되어 있다고 보고하였다. Kim과 Kim(2019a; 2019b)은 동영상과 스피치가 동일한 비중으로 주제를 표현하는 데에 사용된 비디오들을 실험 자료로 사용했을 때 비디오 주제에 적합한 자극(비디오)에 대해서 전측 P600 효과가 나타난다고 보고하였다. 앞의 N400 효과와 동일하게 스피치 중심의 비디오를 자극으로 주었을 때 스피치와 함께 제시되는 동영상

으로 인하여 P600 효과에서도 측두엽 영역의 활성화가 상쇄될 수 있다고 가정하였다. 이에 따라서 연구 가설 2는 다음과 같이 두 개로 세분화된다.

- (1) 연구 가설 2-1: 비디오 실험자료의 주제 적합 쇼트는 전측에서 P600 효과를 나타낼 것이다.
- (2) 연구 가설 2-2: 사운드 실험자료의 주제 적합 쇼트는 후측 및 측두엽에서 P600 효과를 나타낼 것이다.

3. 실험 설계

실험을 수행하기 위해서 피험자들과 실험 자료를 선정하고, 실험 처치와 실험 절차를 설계하였다. 설계된 절차에 따라 실험을 수행한 후에 수집한 뇌파 정보를 분석하여 연구 가설 1과 2를 검증하였다. 실험을 수행하기 전에 명지대학교 기관생명윤리위원회의 승인을 받았다(IRB 승인번호: MJU-2018-02-001-02).

3.1 피험자, 실험 자료 및 실험 처치

뇌파는 나이와 성별에 따라서 차이가 있다고 알려졌다(Evans, Cui, & Starr, 1995). 따라서 본 실험에서는 피험자들로 20대의 오른손잡이 남자로 제한하여 M대학교와 Y대학교의 학부생과 대학원생 36명을 모집하였다(학부생: 34명, 대학원생: 2명). 뇌파 실험의 자료로 매우 짧은 시간에 나타나는 ERP의 N400 및 P600의 효과가 비디오에 의한 것인지 아니면 사운드에

〈표 1〉 실험 자료와 실험 처치

구분	실험 처치의 내용	실험 자료의 명칭	실험 자료의 특성
통제 조건	〈빅데이터의 활용〉 제목의 원 비디오(재생시간: 1분 10초)를 13개 쇼트로 구분 제시	비디오 (실험)자료	스피치 + 동영상 (통합모형 반영)
실험 조건	위 통제조건의 13개 비디오 쇼트들의 동영상상을 제거한 후, 스피치 내용을 화면 자막으로 제시	사운드 (실험)자료	스피치 + 자막 (언어모형 반영)

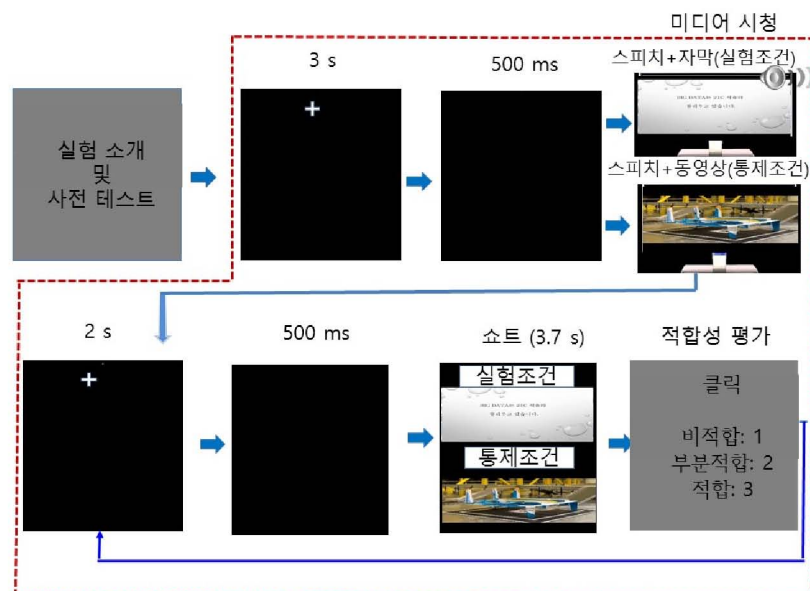
의한 것인지를 조사하기 위해서 두 가지 유형의 미디어를 사용하였다(〈표 1〉 참조).

피험자 36명을 두 개의 그룹으로 무작위로 균등하게 나눈 후 통제 집단(18명)에게 원래의 비디오를 13개 쇼트들로 구분하여 시청시켰다. ERP 실험의 자극 제시전 200ms 동안 배경뇌파를 측정하기 위해 쇼트간 경계는 0.5초 기간 동안의 페이드아웃 장면전환효과를 주었다. 실험 집단(18명)에게는 동영상은 제거한 후 자막으로 대체한 스피치와 자막으로 구성된 총 13개의 쇼트들로 구성된 자료를 시청시켰다. 통

제 집단과 마찬가지로 쇼트간 경계는 0.5초 기간 동안의 장면전환 효과를 주었다.

3.2 실험 절차

피험자들에게 실험 절차를 설명한 후 뇌파기기(캡)을 피험자에게 장착시킨 후 세 개의 쇼트로 구성된 짧은 비디오를 사용하여 사전 테스트를 수행하여 피험자들에게 실험 절차를 숙지시켰다. 이후 세션 1에서는 제1그룹에 속한 피험자들에게 사운드 실험 자료(스피치+자



〈그림 2〉 실험 절차

막)를 시청하게 한 후 사운드에서 추출한 사운드 쇼트들을 순서대로 각각 제시한 후 현재 쇼트를 단서로 하여 방금 전 시청한 사운드의 주제와의 적합도(적합: 3, 부분적합: 2, 부적합: 1)를 평가하도록 하였다. 이때 피험자들에게 각 쇼트의 전체를 다시 시청하게 하는 것 대신 처음 3.7초 동안만 시청하게 하였다. 세션 2에서는 제2그룹에 속한 피험자들에게 비디오 실험자료(스피치+동영상)를 시청하게 한 후 사운드 자료에서와 같은 방식으로 적합성 평가 단계를 거쳤다. 이와 같이 두 개의 세션에서 모두 미디어 시청 단계와 적합성 평가 단계를 따로 구분하는 것은 만약 피험자들에게 현재 시청하고 있는 미디어 내용을 동시에 평가하게 하면 평가 결과를 입력하는 버튼 클릭과 같은 동작에 의해서 행동뇌파(motor EEG)가 발생할 수 있어 연구의 타당성을 저해할 수 있기 때문이다(Luck, 2014).

3.3 뇌파측정과 분석

뇌파를 측정하기 위해서 Neuroscan Synamp amplifier(Compumedics Neuroscan)의 30채널 Quick Cap을 사용하였고, 기준(reference) 전극으로 A1과 A2를 이용하였다. 스피치/동영상 또는 스피치/자막 제시와 뇌파측정의 시작점 동기화를 통제하기 위해서 E-Prime을 이용하였다. 측정된 뇌파자료는 CURRY 7.012(Compumedics Neuroscan) 프로그램을 이용하여 분석하였다. 36명의 피험자들로부터 수집한 뇌파 데이터 중에서 데이터의 품질이 낮거나 뇌파가 제대로 측정되지 못한 4명을 제외시키고 32명(통제집단 16명, 실험 집단 16명)의 데이터를 분석하였

다. 뇌파 측정은 피험자들이 미디어 전체를 시청할 때의 뇌파와 미디어 시청 후 3.7초 동안 제시된 쇼트를 시청할 때의 뇌파와 구분하여 측정하였다. 이 경우 미디어 내의 해당 쇼트의 시작점을 찾아서 -200ms에서 시작점까지를 기저선으로 하고, 시작점에서부터 1,000ms에 대한 뇌파 데이터를 각 미디어 쇼트를 시청할 때 나타나는 뇌파 반응으로 분류하여 분석하였다. 따라서 각 에폭의 길이는 총 1.2초이다.

사운드 자료의 경우, 총 에폭 208개(피험자당 13개 에폭들)는 주제적합 에폭 77개, 주제부분적합 에폭 70개, 주제비적합 에폭 61개로 구분되었다. 연구 가설들을 검증하기 위해서 주제부분적합 에폭들은 애매성 때문에 제외시키고 138개의 에폭들만 이용하였다. 한편 비디오 자료의 경우, 총 에폭 208개 중에서 6개 에폭들은 피험자들이 신체를 심하게 움직여 생기는 잡음으로 인하여 제거시키고 나머지 에폭들은 주제적합 에폭 51개, 주제부분적합 에폭 71개, 주제비적합 에폭 81개로 구분되었다. 앞의 사운드 자료와 동일하게 주제부분적합 에폭들을 제외한 132개의 에폭들만 연구 가설 검증에 이용되었다.

미디어 쇼트를 단서로 하여 방금 전에 시청한 미디어 주제와의 적합도를 평가하는 것은 미디어 시청 단계와 적합성 평가 단계간의 불일치성을 발생시킬 가능성이 있다. 즉, 사운드 자료를 시청하는 동안에는 특정 쇼트가 주제에 적합하다고 생각했지만 시청이 끝난 후 사운드 주제가 명확해지면 해당 쇼트가 주제에 적합하지 않다고 판단할 가능성이 있기 때문이다. 인공신경망을 이용하여 이러한 불일치도가 어느 정도 되는지 알아보았다. 인공신경망의 입력

데이터로 피험자들이 미디어를 시청하는 동안 미디어 쇼트에 반응하는 뇌파 데이터와 피험자들이 미디어 시청 후 해당 쇼트에 부여한 적합도가 초기값으로 사용된다. 분류 결과, 사운드와 비디오 자료는 각각 높은 분류 정확도(83.3%, 93.3%)를 보여주어서 미디어를 시청할 때의 각 쇼트에 대한 뇌파 반응과 시청 후 각 쇼트의 적합도 평가 결과간에 일치도가 매우 높은 것으로 나타났다.

4. 실험 결과

4.1 연구 가설 검정을 위한 통계분석

연구 가설 1과 2의 검증을 위해서 비디오 실험자료(스피치+동영상) 제시조건과 사운드 실험자료(스피치자막) 제시조건에서 수집한 주제 적합 쇼트들과 주제 비적합 쇼트들의 N400 잠재기에 해당하는 기간 동안의 음전위 최저값, P600 잠재기에 해당하는 양전위 최고값을 비교하였다. ERP 분석결과에 대해 '비디오 제시조건'과 '사운드 제시조건'을 개인간 요인(between-subject factor)으로 하고 '주제 적합'과 '주제 비적합'을 개인내 요인(within subject factor)으로 하는 2×2 요인설계에 따른 반복측정(repeated measures) MANOVA 검증을 실시하였다. 상호작용의 구체적 양상을 살펴보기 위해 이외에도 통제조건(스피치+동영상)과 실험조건(스피치+자막)의 개별 상황에서 주제관련성(적합 vs. 비적합)의 효과에 대하여 반복측정 t-검증을 실시하였다(〈표 2〉와 〈표 3〉의 비디오와 사운드의 t값 참조).

4.2 연구가설 1의 검정 결과

N400 효과에서 비디오 자료가 사용된 경우 전측 N400 효과를 나타낼 것으로 예측하였다. 사운드 자료의 경우 후측 N400 효과를 나타낼 것이며 측두엽 영역도 함께 활성화될 것으로 가정하였다. 뇌파 실험에서 측정한 미디어별 주제 적합쇼트와 주제 비적합쇼트에 대한 N400 구간의 음전위 최저점의 평균값을 비교한 결과는 다음과 같다. 비디오의 발현 범위는 예측한 대로 전두엽(F8)과 전두-중앙엽(FC4)의 두뇌의 전측으로 나타났다. 한편 예측과 다르게 측두엽 영역(TP8)도 활성화되는 것으로 나타나 연구 가설 1-1은 부분적으로 검증되었다. 한편 사운드 자료의 경우 전두엽(F8)과 측두엽(TP8, FT8)에서만 N400 효과가 나타났다. 즉, 후측 N400 효과는 없었으나 측두엽 영역은 예측한 대로 활성화되었다. 이에 따라서 연구 가설 1-2도 부분적으로 검증되었다.

주제 관련성의 주효과는 전전두엽(Fp2), 전두엽(F4, F8), 중앙엽(C4), 측두엽(T8, TP8, FT8), 전두-중앙엽(FC4)의 8개의 채널에서 나타났다. 앞에서 확인한 것처럼 비디오와 사운드 모두에서 전측 및 측두엽 영역에서 N400 효과가 나타났기 때문에 미디어의 주효과와 미디어와 주제 관련성의 상호 작용 효과는 나타나지 않았다.

4.3 연구가설 2의 검정 결과

P600 효과에서 비디오 자료의 경우 전측 P600 효과를 나타낼 것으로 예측하였다. 한편 사운드 자료가 사용된 경우 후측 P600 효과를

〈표 2〉 N400의 비디오 자료(스피치 + 동영상)와 사운드 자료(스피치 + 자막)의 주효과 및 상호작용 효과

영역	채널	비디오 자료 (N=16)			사운드 자료 (N=16)			주효과 (F값)		상호작용 효과(F값)
		비적합 쇼트	적합 쇼트	t값	비적합 쇼트	적합 쇼트	t값	미디어 (사운드 vs. 비디오)	주제 관련성(적합 vs. 비적합)	미디어× 주제관련성
전전두엽	FP1	-13.25	-8.03	1.15	-12.56	-12.15	0.01	0.31	0.77	0.56
	FP2	-23.27	-12.96	3.58	-24.09	-15.77	2.84	0.20	6.42*	0.07
전두엽	Fz	-12.59	-9.79	1.45	-9.79	-8.13	0.22	0.65	1.10	0.07
	F3	-5.86	-5.81	0.00	-5.76	-6.41	0.02	0.00	0.01	0.01
	F4	-15.84	-9.96	2.40	-15.34	-8.74	3.54	0.06	5.84*	0.01
	F7	3.29	-5.62	4.26	-5.08	-8.44	0.46	2.65	3.52	0.72
	F8	-28.33	-16.42	8.00*	-25.96	-16.13	7.91*	0.08	15.79***	0.14
측두엽	T7	-1.99	-3.50	0.39	-7.28	-5.64	0.13	1.68	0.00	0.39
	T8	-11.80	-6.65	4.41	-11.31	-7.74	3.82	0.10	8.14**	0.26
	FT7	-0.76	-4.87	1.49	-3.99	-6.02	0.28	0.74	1.46	0.16
	FT8	-16.43	-10.32	2.61	-15.78	-8.99	8.45*	0.10	8.43**	0.02
	TP7	-1.84	-2.62	0.05	-5.19	-4.06	0.11	0.31	0.00	0.16
	TP8	-10.47	-6.52	4.97*	-8.32	-2.12	8.52*	1.89	13.47***	0.67
중앙엽	Cz	-9.48	-7.56	1.62	-7.40	-7.05	0.02	0.35	0.64	0.30
	C3	-4.81	-5.81	0.25	-6.53	-6.36	0.00	0.26	0.05	0.10
	C4	-11.40	-9.43	1.25	-9.90	-6.40	1.74	1.06	2.95	0.23
전두- 중앙엽	FCz	-9.58	-8.08	0.67	-8.75	-7.78	0.09	0.06	0.44	0.02
	FC3	-5.70	-6.23	0.06	-6.62	-6.75	0.00	0.09	0.02	0.00
	FC4	-12.97	-8.15	5.48*	-12.55	-7.09	3.39	0.12	8.12**	0.03
중앙- 두정엽	CPz	-8.95	-7.39	1.28	-6.83	-6.34	0.04	0.53	0.58	0.16
	CP3	-5.94	-7.02	0.53	-6.04	-6.05	0.00	0.04	0.12	0.11
	CP4	-10.17	-7.88	2.13	-8.15	-5.49	1.06	0.94	2.68	0.01
두정엽	Pz	-9.36	-7.98	0.69	-6.29	-6.23	0.00	0.90	0.28	0.24
	P3	-6.59	-7.02	0.08	-6.50	-5.76	0.07	0.12	0.01	0.14
	P4	-8.53	-7.17	0.76	-6.48	-4.85	0.45	1.15	1.08	0.00
	P7	-2.60	-4.11	1.06	-4.22	-2.64	0.29	0.00	0.00	0.90
	P8	-7.16	-6.95	0.01	-7.38	-3.60	2.12	0.27	1.76	1.42
후두엽	Oz	-6.91	-6.09	0.20	-6.81	-3.72	1.90	0.34	1.84	0.62
	O1	-6.39	-5.32	0.55	-7.06	-4.72	1.07	0.00	1.64	0.22
	O2	-6.19	-8.47	1.60	-5.72	-4.45	0.34	0.47	0.12	1.58

N400: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0017$ (본페로니 교정)

나타낼 것이며 측두엽 영역도 활성화될 것으로 가정하였다. 미디어별 주제 적합쇼트와 주제 비적합쇼트에 대한 P600 구간의 양전위 최고 점의 평균값을 비교한 〈표 3〉을 살펴보면 다음

과 같다. 비디오의 경우 발현 범위가 전전두엽 (FP2), 전두엽(F8), 전두-중앙엽(FC4), 중앙-두정엽(CP4), 두정엽(P4)으로 나타나 예측한 것보다 더 넓게 나타났다. 측두엽(T8, FT8,

〈표 3〉 P600의 비디오 자료(스피치 + 동영상)와 사운드 자료(스피치 + 자막)의 주효과 및 상호작용 효과

영역	채널	비디오 자료 (N=16)			사운드 자료 (N=16)			주효과 (F값)		상호작용 효과(F값)
		비적합 쇼트	적합 쇼트	t값	비적합 쇼트	적합 쇼트	t값	미디어 (사운드 vs. 비디오)	주제 관련성(적합 vs. 비적합)	미디어× 주제관련성
전전두엽	Fp1	-1.78	10.41	3.90	1.35	9.10	2.10	0.06	5.97*	0.29
	Fp2	-8.32	8.70	6.80*	2.77	13.99	4.53	2.82	11.33**	0.48
전두엽	Fz	-1.52	1.88	1.65	3.67	10.69	2.29	5.79*	3.81	0.45
	F3	4.68	8.87	0.84	4.47	9.31	1.16	0.00	1.99	0.01
	F4	-4.73	2.97	3.88	3.47	12.54	5.89*	6.79*	9.63**	0.06
	F7	17.15	15.65	0.08	10.07	7.30	0.19	4.30*	0.27	0.02
	F8	-14.03	2.64	19.38***	4.60	17.72	12.87**	11.91**	32.02***	0.45
	F8	-14.03	2.64	19.38***	4.60	17.72	12.87**	11.91**	32.02***	0.45
측두엽	T7	9.98	10.63	0.04	7.00	11.35	0.61	0.17	0.63	0.34
	T8	-0.83	6.01	6.38*	5.54	12.34	12.94**	4.28*	19.12***	0.01
	Ft7	11.49	11.46	0.00	8.63	8.21	0.01	1.31	0.00	0.00
	Ft8	-5.63	2.88	6.42*	5.27	16.26	7.87*	14.16***	14.29***	0.22
	TP7	8.00	9.76	0.19	6.01	10.89	1.97	0.00	1.59	0.35
	TP8	-1.18	2.54	8.35*	5.31	15.73	8.42*	12.11**	13.75***	3.06
중앙엽	Cz	2.90	4.21	0.66	4.41	11.49	3.72	3.10	4.38*	2.07
	C3	5.93	7.33	0.35	5.50	10.09	1.49	0.27	1.82	0.51
	C4	-0.23	1.64	1.38	4.66	12.23	4.88*	9.69**	6.24*	2.26
전두- 중앙엽	FCz	2.60	4.55	0.85	4.28	11.29	3.07	3.19	3.93	1.25
	FC3	5.74	6.44	0.07	5.75	9.91	1.08	0.57	1.03	0.52
	FC4	-1.74	3.01	7.03*	4.14	12.92	8.96**	12.13**	15.52***	1.36
중앙- 두정엽	CPz	3.19	4.40	1.05	4.52	10.72	2.94	2.43	3.79	1.72
	CP3	4.92	5.95	0.48	5.13	9.86	1.90	0.81	2.37	0.97
	CP4	1.37	3.99	7.94*	5.45	11.77	3.62	5.80*	6.71*	1.16
두정엽	Pz	2.48	4.43	2.66	4.70	10.39	2.69	2.40	4.34*	1.04
	P3	4.68	4.25	0.11	5.10	9.89	1.98	1.73	1.44	2.06
	P4	2.21	3.61	4.93*	4.37	10.88	3.68	5.46*	5.26	2.19
	P7	6.15	5.72	0.05	4.14	9.88	2.37	0.13	1.60	2.17
	P8	1.25	1.21	0.00	3.23	11.65	4.91*	3.71	4.38*	4.47*
후두엽	Oz	3.64	4.10	0.09	4.48	10.18	2.93	1.67	2.84	2.05
	O1	4.08	5.24	0.53	4.32	9.91	2.55	0.58	3.08	1.33
	O2	4.47	2.21	1.66	4.66	9.88	2.25	1.27	0.57	3.68

P600: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0017$ (본페로니 교정)

TP8) 영역도 예측과 다르게 P600 효과를 나타냈다. 따라서 연구 가설 2-1은 검증되지 못했다.

한편 사운드 자료에서 전두엽(F4, F8), 중앙엽(C4), 전두-중앙엽(FC4), 두정엽(P8)에서

P600 효과를 나타내 예측한 것과 달리 발현 범위가 더 넓게 나타났다. 또한 측두엽 영역(T8, FT8, TP8)에서 높은 양전위값들이 나타났다. 특히 T8에서 유의 수준 0.01에서 적합 사운드

의 양전위값(12.34)이 비적합 사운드의 양전위값(5.54)보다 통계적으로 유의미하게 높게 나타났다. 따라서 연구 가설 2-2는 부분적으로 검증되었다.

미디어의 주효과는 전두엽(Fz, F4, F7, F8), 중앙엽(C4), 전두-중앙엽(FC4), 측두엽(T8, TP8, FT8), 중앙-두정엽(CP4), 두정엽(P4)의 11개의 채널에서 나타났다. 한편 주제 관련성의 주효과는 전전두엽(Fp1, Fp2), 전두엽(F4, F8), 중앙엽(Cz, C4), 전두-중앙엽(FC4), 측두엽(T8, FT8, TP8), 중앙-두정엽(CP4), 두정엽(Pz, P8)의 13개의 채널에서 나타났다. 미디어와 주제 관련성의 상호 작용 효과는 P8에서 나타났다. 이는 사운드에서 P8의 P600 효과가 나타났으나 비디오에서는 P8의 P600 효과가 나타나지 않았다는 것을 의미한다.

5. 논의와 활용

5.1 논의

5.1.1 전후측 N400 효과

N400 효과에서 비디오의 경우 전두엽, 전두-중앙엽, 측두엽의 뇌파활성도 세기에 의해서 주제에 적합하지 않는 쇼트를 가려낼 수 있음을 확인하였다. 예측한 것과 달리 측두엽 영역도 활성화되는 것으로 나타났다. 이는 실험 데이터로 사용한 비디오가 스피치에 의해서 주로 정보를 제공하고 보조적으로 동영상을 사용하는 자료이기 때문으로 생각된다. 특히 일부 동영상 쇼트들은 스피치 내용을 단어 또는 문장으로 요약하여 화면 자막으로 제시하는 경우도 있다. 이

는 만약 비디오내의 동영상이 많은 정보를 제공하는 미디어를 실험 자료로 사용할 경우에는 그 결과가 달라질 수 있음을 의미한다.

사운드 자료에 비하여 비디오 자료의 주제 비적합 쇼트의 N400 음전위 최저값이 주제 적합 쇼트의 최저값보다 낮아지는 경향을 보이고 있다. 이러한 결과가 나온 이유들을 유추해 보면 다음과 같다. 먼저 제시되는 일부 비디오 쇼트의 동영상 정보가 비적합 쇼트를 가려내는 데 어떤 역할을 하지 않았나 추측된다. 또한 사운드 자료에서는 자막이 스피치의 의미와 일치하고 의미를 상호 강화하는 반면 비디오 자료에서는 동영상이 전하는 의미정보가 스피치와 일치하지 않을 수도 있기 때문으로 생각된다.

N400에서 중요한 역할을 하는 채널(들)을 파악하기 위해서 판별분석을 수행하였다. 판별 분석을 위해서 적합 쇼트와 비적합 쇼트 간에 통계적으로 유의미한 차이가 나는 최소 전위값들의 평균들(N400)을 종속 변수들로 사용하였다. 분석 결과, F8(전두엽)이 비디오 쇼트의 주제 비적합성 여부를 가려내는 데 가장 중요한 판별 변수로 나타났다.

사운드 자료의 경우는 전두엽과 측두엽의 뇌파활성도 세기에 의해서 주제에 적합하지 않는 쇼트를 가려낼 수 있음을 확인하였다. 또한 언어가 문법적이나 의미적으로 오류가 있을 때 나타나는 효과로 알려진 후측 N400 효과가 약하게 나타나는 이유는 실험 자료로 사용한 스피치와 자막의 내용이 문법적이나 의미적으로 오류가 없을 뿐만 아니라 스피치와 자막의 의미가 서로 일치하기 때문으로 생각된다. 판별분석 결과, TP8(측두-두정엽)이 주제 적합성 여부를 가려내는 데 가장 중요한 변수로 나타났다.

5.1.2 전후측 P600 효과

비디오의 경우 P600 효과가 전두엽, 전전두엽, 전두-중앙엽, 측두엽, 중앙-두정엽, 두정엽에서 나타났다. 예측한 것과 달리 측두엽 영역도 활성화되는 것은 앞에서 언급한 것처럼 실험에 사용된 비디오가 스피치에 의해서 주로 정보를 제공하고 있기 때문으로 생각된다. 판별 분석을 위해서 두 종류의 쇼트간에 통계적으로 유의미한 차이가 나는 최대 전위값들의 평균들(P600)을 종속 변수들로 사용하였다. F8(전두엽)이 주제 적합성을 판정하는 데 가장 중요한 판별 변수로 나타났다.

한편 사운드에서는 P600 효과가 전두엽, 중앙엽, 전두-중앙엽, 측두엽, 두정엽에서 나타났다. 특히 T8(측두엽)에서 유의 수준 0.01에서 적합 사운드의 양전위값이 비적합 사운드의 양전위값보다 통계적으로 유의미하게 높게 나타났다. 비디오의 경우와 동일하게 사운드에서도 F8(전두엽)이 주제 적합성을 판정하는 데 가장 중요한 판별 변수로 나타났다.

사운드가 비디오에 비해서 P600 효과를 더 뚜렷하게 나타낸 이유는 자극으로 스피치와 자막을 함께 사용했기 때문으로 추측된다. 다시 말해서 스피치 제시 이후 500ms~700ms(0.5초~0.7초) 사이의 두뇌 활동을 관찰하는 P600의 경우 0.7초 이내에 시청한 정보는 단어 하나 또는 많아야 두 단어들을 듣지만 동시에 자막의 정보를 스캔할 수 있어서 스피치와 자막이 서로 보완 효과가 생겨나 해당 사운드 쇼트의 의미 파악이 좀 더 용이했기 때문으로 생각된다. 만약 자막 없이 휘발성이 강한 스피치만 제시한다면 P600 효과의 강도는 달라질 수 있을 것이다. 이에 반해서 스피치에 중점을 두고 동

영상 자료를 보조 자료로 활용한 비디오의 경우 0.7초 동안 정지 화상 또는 동영상상을 보면서 단어 하나 또는 두 개를 듣는 피험자가 해당 비디오 쇼트의 전체 내용을 파악하기에 정보가 충분하지 않을 수도 있어서 약한 P600 효과를 냈다고 추측된다.

요약하면 본 연구에서 얻어진 N400과 P600의 효과에 대한 결과는 뇌파 실험 자료로 사용된 언어 메시지에 주로 의존하는 다큐멘터리 비디오의 특성에서 비롯된다고 보인다. 따라서 다른 유형의 비디오 자료를 실험에 사용한다면 그 결과가 달라질 수 있을 것이다.

5.2 활용 방안

본 연구 결과는 사운드 요약은 물론 비디오 스킵을 생성하는 데에 활용할 수 있다. 일반적으로 메타데이터 요소들은 개인의 취향에 맞추기보다는 표준화된 메타데이터 요소들을 구성하는 것이 일반적이다. 그러나 최근 개인이 시청한 미디어를 나중에 다시 보거나 재활용하기 위해서 특정 개인이 관심을 갖고 시청한 쇼트들을 추출하여 구성한 개인화된 미디어 요약에 대한 요구도 생겨나기 시작했다(Moon, Kwon, Park, & Yoon, 2019). 이러한 일반적인 또는 개인화된 내용 기반 메타데이터를 생성하기 위해서는 다음과 같은 작업이 필요하다.

첫째, 일반적인 사운드 요약(또는 비디오 스킵)을 생성하기 위해서 기존 피험자들이 미디어를 시청할 때 수집된 뇌파 정보가 새로운 미디어를 분석하기 위한 훈련 데이터로 사용될 수 있다. 이후 비교적 표준화된 뇌파 패턴을 갖는 피험자들에게 뇌파기기를 장착시킨 후 요약 대상

의 새로운 미디어를 시청하게 한다. 이때 본 연구 결과를 분류 알고리즘(예, 판별 분석의 판별 함수나 인공신경망) 설계에 활용하여 주제에 적합한 미디어 쇼트를 추출하여 미디어 요약을 구성할 수 있다. 예를 들어서, 특정 사운드 쇼트의 ERP 데이터가 전두엽(F8)에서 강한 P600 효과를 나타내는 경우 이 쇼트를 추출하여 사운드 요약을 생성할 수 있을 것이다. 궁극적으로 뇌파 정보를 활용하여 구성된 사운드 요약(또는 비디오 스킴)들과 원자료들간의 관계를 딥러닝(deep learning) 방식으로 분석하여 새로운 미디어의 요약을 자동으로 생성할 수 있는 시스템을 구현할 수 있을 것으로 기대해 본다.

둘째, 개인화된 맞춤형 사운드 요약(또는 비디오 스킴)을 생성하기 위해서 특정 개인이 미디어를 시청하는 동안 생성되는 뇌파 정보가 필요하다. 이러한 뇌파 정보가 없는 경우에는 사전 테스트를 통해서 미디어 주제에 적합한 또는 적합하지 않은 쇼트들을 시청할 때 생기는 뇌파 정보를 수집한다. 이후 이러한 사전 테스트를 통해서 얻은 특정 개인의 뇌파 데이터만을 훈련 데이터로 사용하거나 또는 이 뇌파 데이터를 기존 피험자들의 뇌파 데이터와 결합한 후 이를 훈련 데이터로 사용하여 개인화된 맞춤형 미디어 요약을 자동 생성할 수 있을 것이다.

6. 결 론

본 연구는 인지적 정보처리 모형을 구성하고 이 모형을 뇌파측정기술에 접목시켜 사운드 요약을 하는 방안들을 제안하기 위해서 연구 가설들을 수립하고 검증해 보았다. 검증 결과를

요약하면 다음과 같다.

첫째, N400 효과에서 비디오는 예측한 대로 전측 N400 효과를 나타냈지만 측두엽 영역에서도 N400 효과를 나타냈다. 이는 실험에 사용된 비디오가 스피치에 의해서 주로 정보를 제공하고 보조적으로 동영상 자료를 활용하였기 때문에 비디오 속의 동영상 자료에 대한 피험자들의 뇌파 반응이 크지 않았다고 생각할 수 있다. 한편 사운드에서 약한 후측 N400 효과를 나타낸 것은 스피치의 내용이 문법적이나 의미적으로 큰 오류가 없었기 때문으로 생각된다.

둘째, 비디오의 P600 효과가 후두엽을 제외한 모든 영역들에 걸쳐서 나타났다. 특히 측두엽 영역도 함께 활성화되는 것은 실험에 사용된 비디오가 스피치에 의해서 주로 정보를 제공할 뿐 아니라 일부 동영상 쇼트들은 스피치 내용을 화면 자막으로 제시하는 경우도 있기 때문으로 생각된다. 한편 사운드에서는 P600 효과가 전전두엽과 후두엽을 제외한 영역들에 걸쳐서 나타났다. 사운드가 비디오에 비해서 P600 효과를 더 뚜렷하게 나타내는 이유는 휘발성이 강한 스피치와 함께 자막을 사용하여 이 두 정보가 상호보완 되었기 때문으로 추측된다.

이러한 결과에 기초하여 다음과 같은 후속 연구들을 제안해 본다.

첫째, 특별히 모순된 내용을 포함하지 않는 사운드자료에서는 N400 효과 없이 P600 효과가 바로 나타날 수 있는 가능성이 높아 보인다. 왜냐하면 기존의 언어 관련 뇌파 실험에서는 의도적으로 문장의 내용을 문법적 또는 의미적으로 왜곡 시킨 후 실험을 하는 사례들을 통해서 N400 효과를 확인했기 때문이다. 더욱이 이런 경우에서조차 N400 효과 없이 P600 효과

가 나타난다고 보고하는 연구가 있다(Kim & Osterhout, 2005). 즉, 사람들은 제시되는 문장이 문법적으로 오류가 있다 해도 문장의 시맨틱 통합에 더 집중하기 때문에 이러한 문법적 오류에 상대적으로 덜 관심을 둔다는 것이다.

P600 효과를 좀 더 효율적으로 분석하기 위해서 각 사운드 쇼트에서 하나 또는 복수개의 자극들이 발생할 수 있다는 점을 고려하여 첫 번째 자극뿐만 아니라 존재한다면 그 다음에 이어지는 자극(들)에 대한 분석도 필요해 보인다. 예를 들어서, 각 쇼트의 첫 번째 자극에 대한 반응과 함께 두 번째 자극에 대한 뇌파 반응도 분석해 볼 수 있을 것이다. 이를 위해서 각 쇼트에서 ERP 반응들이 측정되는 시간 구간을 1초보다 더 길게 잡고 각 쇼트에서 최초의 자극에 의해서 첫 번째 ERP 반응이 나타난다고 가정한다. 이어서 그 다음 자극에 의해서 두

번째 ERP 반응이 발생한다고 가정해 볼 수 있다. 사운드에서 단어가 자극이 될 수 있으며, 이러한 자극의 시작점은 사운드 내용을 분석하여 수집할 수 있을 것이다.

둘째, 비디오에서 각 채널의 특성과 이들 간의 상호작용 효과를 좀 더 정확하게 분석하기 위해서 다음과 같은 후속 연구가 필요해 보인다. 먼저 스피치와 동영상이 거의 유사한 비율로 비디오 내용 및 주제를 전달하는 비디오를 뇌파 실험 자료로 선정한다. 이후 이 비디오를 통합 모형(원 비디오), 언어 모형(자막 없이 스피치만 사용), 영상 모형(동영상)으로 구분하여 세 종류의 미디어들로 구성한다. 피험자들은 세 그룹으로 나뉘고 각 집단에서 각기 다른 미디어를 보도록 한 후 이 세 집단간의 뇌파 반응을 서로 비교해 보는 연구를 수행해 볼 수 있을 것이다.

참 고 문 헌

- Alwehaibi, H. (2015). The impact of using youtube in EFL classroom on enhancing EFL students' content learning. *Journal of College Teaching and Learning*, 12(2), 121-126.
<https://doi.org/10.19030/tlc.v12i2.9182>
- Baddeley, A. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780198528012.001.0001>
- Burmester, J., Spalek, K., & Wartenburger, I. (2014). Context updating during sentence comprehension: The effect of aboutness topic. *Brain and Language*, 137, 62-76.
<https://doi.org/10.1016/j.bandl.2014.08.001>
- Buzzetto-More, N. (2015). Student attitudes towards the integration of youtube in online, hybrid, and web-assisted courses: An examination of the impact of course modality on perception. *MERLOT Journal of Online Learning and Teaching*, 11, 55-73.

- DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, 61, 150-162. <https://doi.org/10.1016/j.neuropsychologia.2014.06.016>
- Evans, W. J., Cui, L., & Starr, A. (1995). Olfactory event-related potentials in normal human subjects: Effects of age and gender. *Electroencephalography and Clinical Neurophysiology*, 95(4), 293-301. [https://doi.org/10.1016/0013-4694\(95\)00055-4](https://doi.org/10.1016/0013-4694(95)00055-4)
- Geyer, A., Holcomb, P., Kuperberg, G., & Perlmuter, N. (2006). Plausibility and sentence comprehension. An ERP study. *Cognitive Neuroscience Supplement*, Abstract, 1-1.
- Hakoda, Y. (2010). *Cognitive psychology: Brain, modeling and evidence*. 강윤봉 (역). (2014). *인지심리학*. 서울: 한국뇌기반교육연구소.
- Hu, W., Xie, N., Li, L., Zeng, X., & Maybank, S. (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. <https://doi.org/10.1109/TSMCC.2011.2109710>
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52(2), 205-225. <https://doi.org/10.1016/j.jml.2004.10.002>
- Kim, H. H., & Kim, Y. H. (2016). Generic speech summarization of transcribed lecture videos: Using tags and their semantic relations. *Journal of the Association for Information Science and Technology*, 67(2), 366-379. <https://doi.org/10.1002/asi.23391>
- Kim, H. H., & Kim, Y. H. (2019a). Video summarization using event-related potential responses to shot boundaries in real-time video watching. *Journal of the Association for Information Science and Technology*, 70(2), 164-175. <http://doi.org/10.1002/asi.24103>
- Kim, H. H., & Kim, Y. H. (2019b). ERP/MMR algorithm for classifying topic-relevant and topic-irrelevant visual shots of documentary videos. *Journal of the Association for Information Science and Technology*, 70(9), 931-941. <https://doi.org/10.1002/asi.24179>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621-647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Maskey, S., & Hirschberg, J. (2006). Summarizing speech without text using hidden markov models. In *Proceedings of the Human Language Technology Conference of the NAACL*

- (Companion Volume: Short Papers, pp. 89-92). Association for Computational Linguistics, Stroudsburg, PA, USA. <https://doi.org/10.3115/1614049.1614072>
- Martin, D. (2018). YouTube: The ultimate 2018 guide to grow your youtube channel, make money fast with proven techniques and foolproof step by step strategies. Cambridge: CreateSpace Independent Publishing Platform.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. The Cambridge handbook of multimedia learning (pp. 134-146). New York: Cambridge University Press.
- Moon, J., Kwon, Y., Park, J., & Yoon, W. C. (2019). Detecting user attention to video segments using interval EEG features. *Expert Systems with Applications*, 115, 578-592. <https://doi.org/10.1016/j.eswa.2018.08.016>
- Nakano, H., Rosario, M. A. M., Oshima-Takane, Y., Pierce, L., & Tate, S. G. (2014). Electrophysiological response to omitted stimulus in sentence processing. *Neuroreport*, 25(14), 1169-1174. <https://doi.org/10.1097/WNR.0000000000000250>
- Nieuwland, M. S., & Martin, A. E. (2012). If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition*, 122(1), 102-109. <https://doi.org/10.1016/j.cognition.2011.09.001>
- van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657-671. <https://doi.org/10.1162/089892999563724>
- Wang, L., & Schumacher, P. B. (2013). New is not always costly: Evidence from online processing of topic and contrast in Japanese. *Frontiers in Psychology*, 4, 363. <https://doi.org/10.3389/fpsyg.2013.00363>
- Wilson, S. M., Bautista, A., & McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *NeuroImage*, 171, 62-74. <https://doi.org/10.1016/j.neuroimage.2017.12.068>
- Zhang, Z., & Fung, P. (2012). Active learning with semi-automatic annotation for extractive speech summarization. *ACM Transactions on Speech and Language Processing*, 8(4), 1-25. <https://doi.org/10.1145/2093153.2093155>