



2025.02.26 (수)

StressID

Multimodal Dataset

Multimodal Dataset for Stress Identification

김동환

목차

Table of Contents

데이터 개요	02p
전반적 개요	
실험 프로토콜	03p
StressID가 수집된 과정	
EDA	21p
결측 모달리티 비율	
연구 방향 제안	23p
Missing Modalities	
Multitask	

데이터 개요

Abstract

StressID

참가자 : 총 65명, 다양한 연령 및 배경(대학원생, 직장인 등)

모달리티 : 생리 신호(ECG, EDA, 호흡), 비디오/오디오

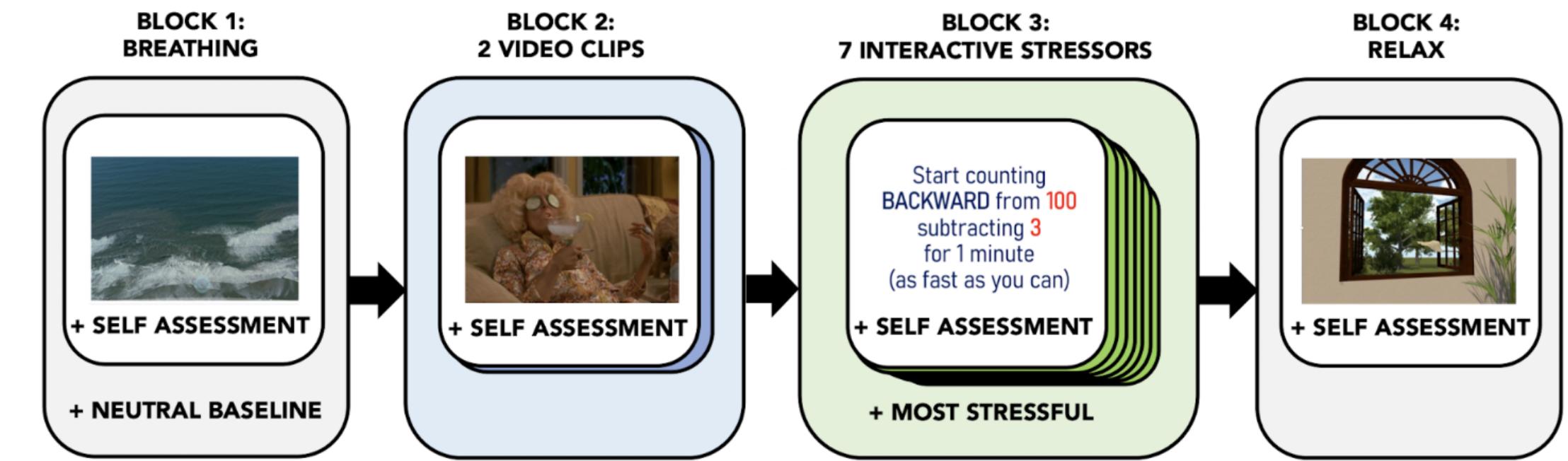
TASK : (Counting, Stroop, Surya, Relax 등)을 통해 다양한 스트레스 유발 및 이완 상태를 기록 이후 자가 평가(스트레스, 이완, 정서, 각성)를 통해 라벨링.



STRESS ID DATASET

실험 프로토콜

Protocol



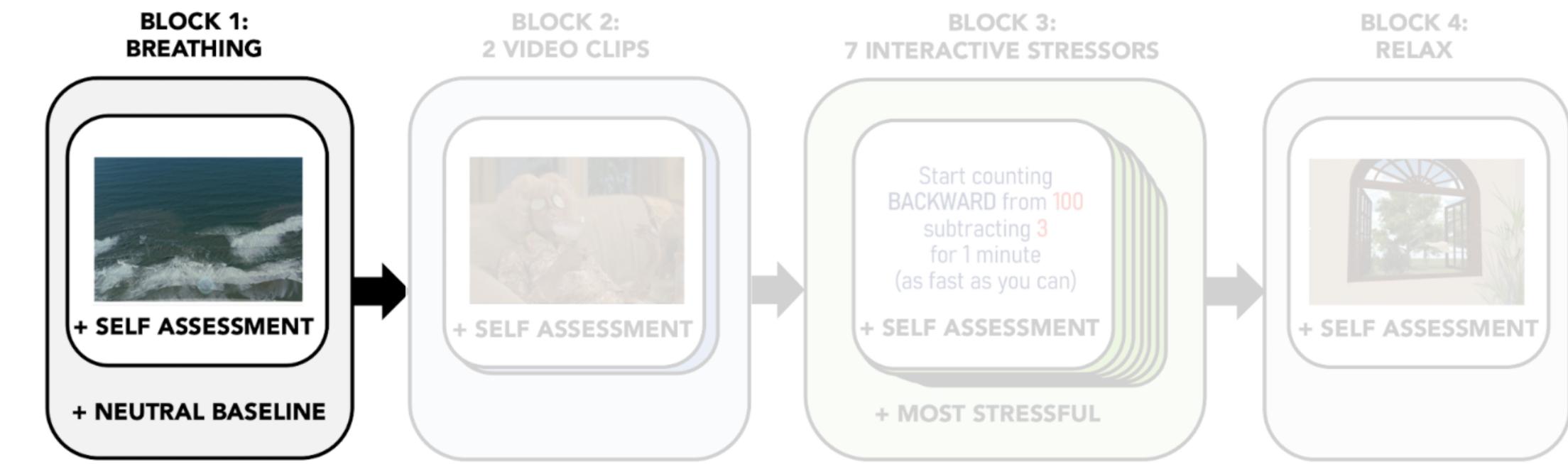
Overview

총 4개의 Block으로 구성된 11개의 TASK

- 1 guided breathing task
- 2 emotional video clips
- 7 interactive stressors
- 1 relaxation task

실험 프로토콜 - 1

Protocol



Block 1 : Breathing

Guided breathing

프로토콜의 첫 번째 블록은 호흡(Breathing) 단일 과제로 구성

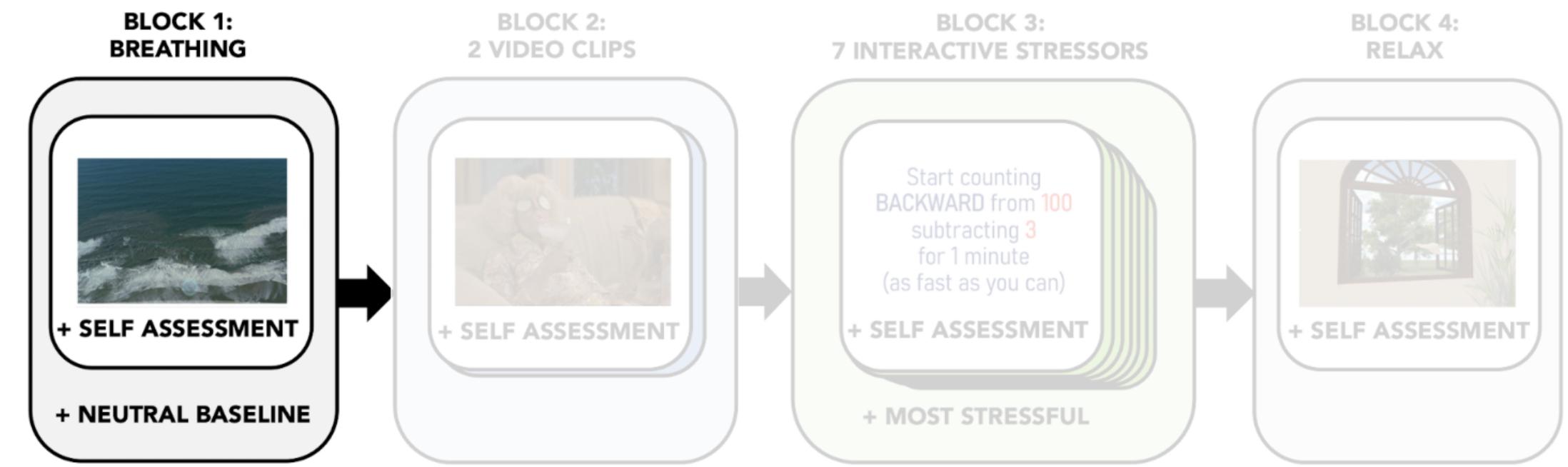
참가자들은 3분 동안 유도 호흡 영상을 시청

참가자들의 정서 상태를 이완하고 중립 상태로 되돌리는 것이 목표

이렇게 기록된 데이터는 각 참가자의 비언어적 중립 상태에 대한 baseline으로 활용

실험 프로토콜 - 1

Protocol

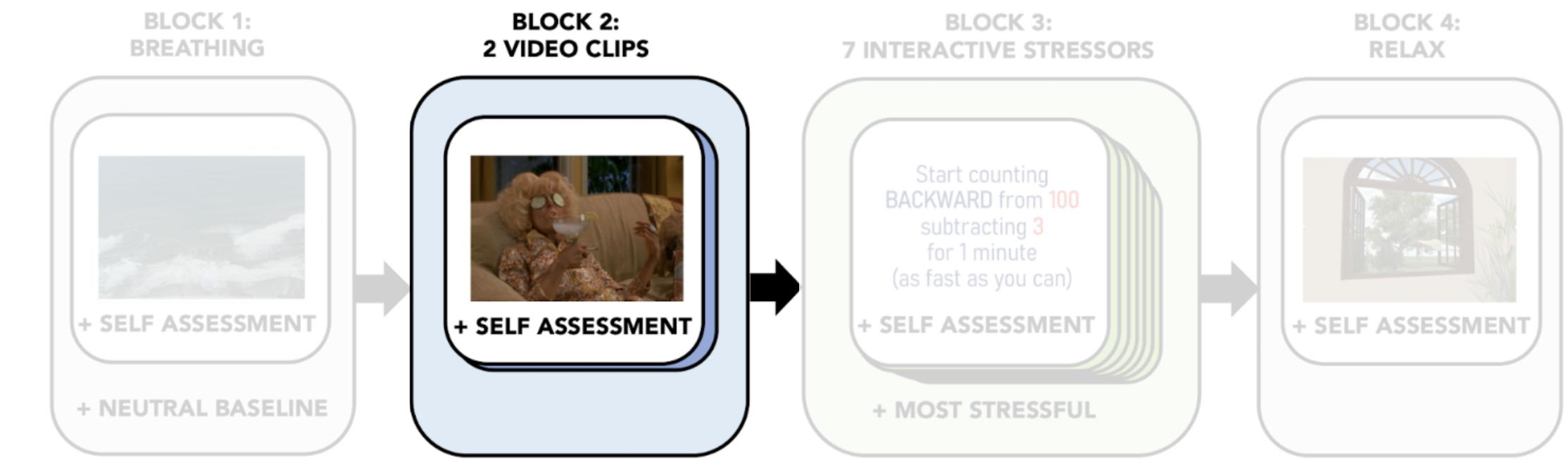


Block 1:
Breathing



실험 프로토콜 - 2

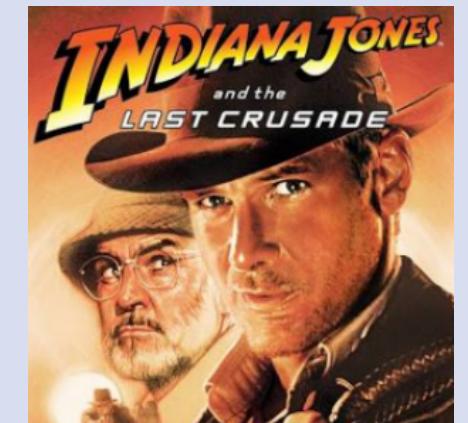
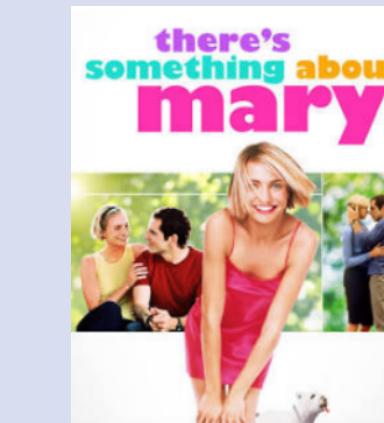
Protocol



Block 2 : Video-clips

Emotional video-clips

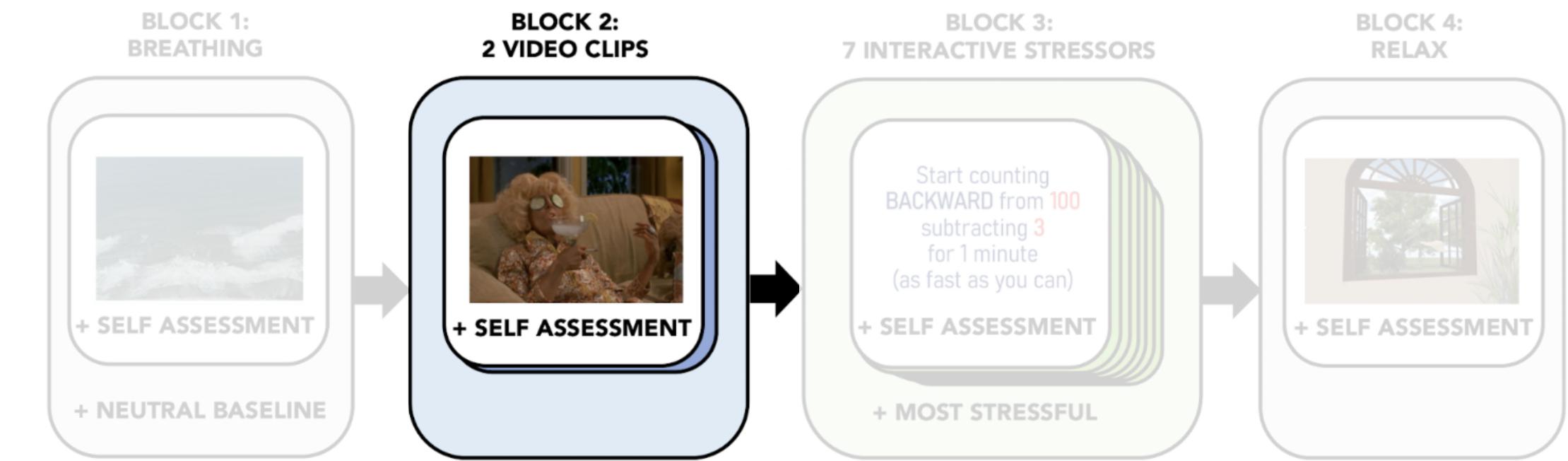
2개의 감정적 비디오 클립을 시청하는 것으로 구성
이 비디오들은 특정 감정적 반응을 유도하기 위해 선정되었음



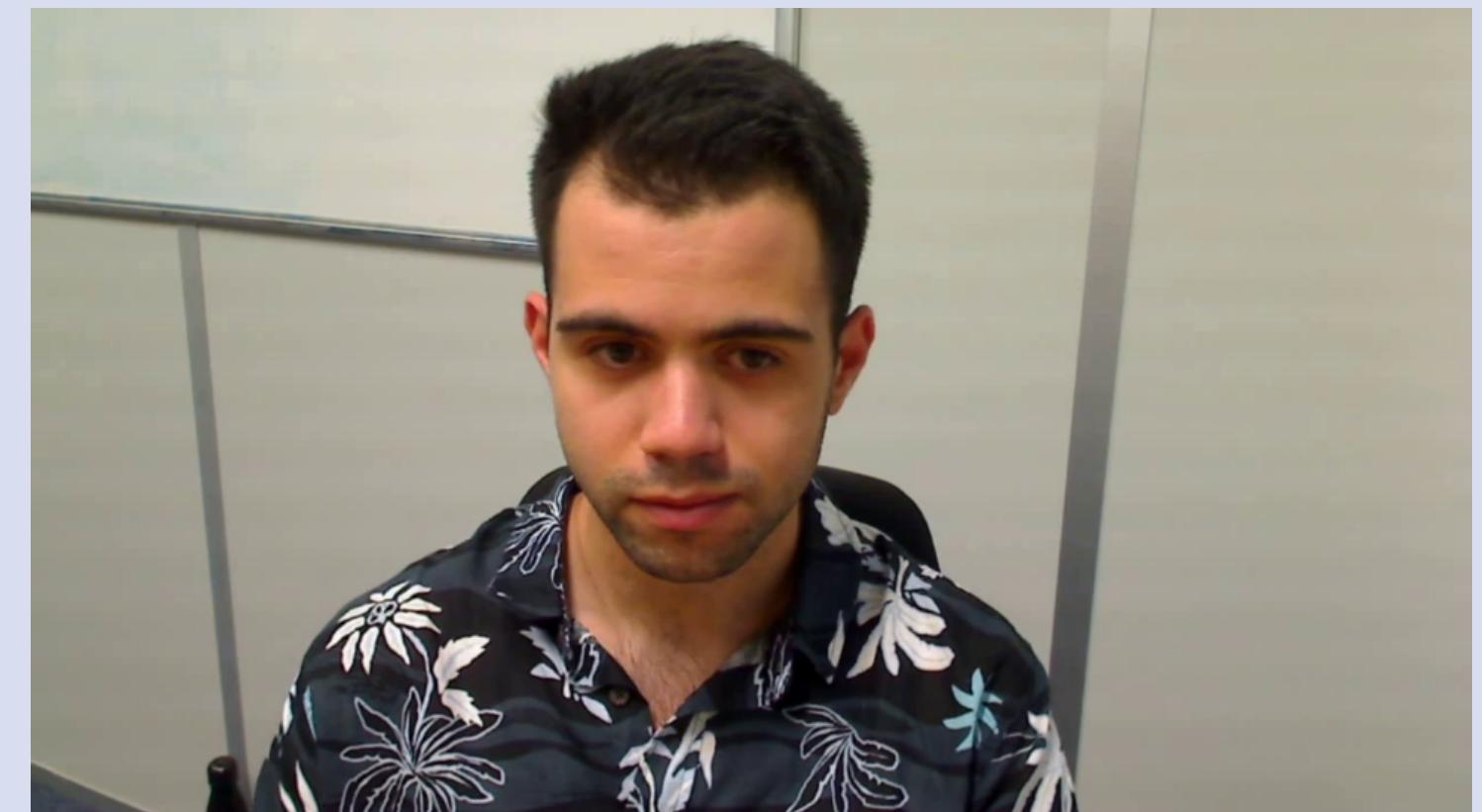
Video1 : 낮은 각성과 긍정적 정서를 유도하기 위해 선정
Video2 : 높은 각성과 부정적 정서를 유도하기 위해 선정

실험 프로토콜 - 2

Protocol

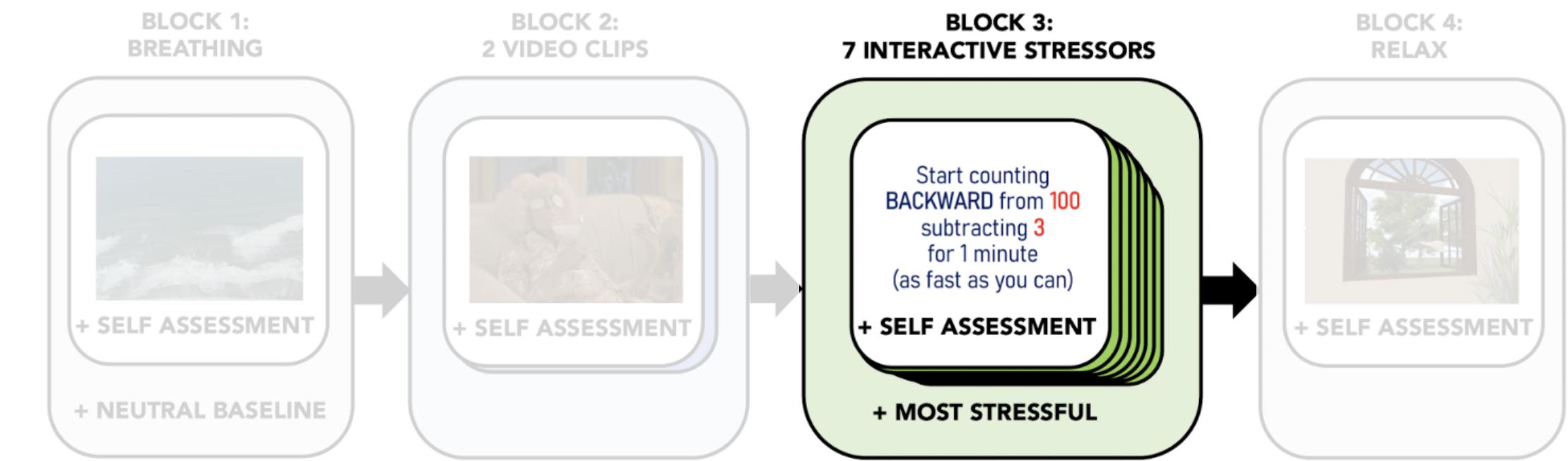


Block 2 :
Video-clips



실험 프로토콜 - 3

Protocol



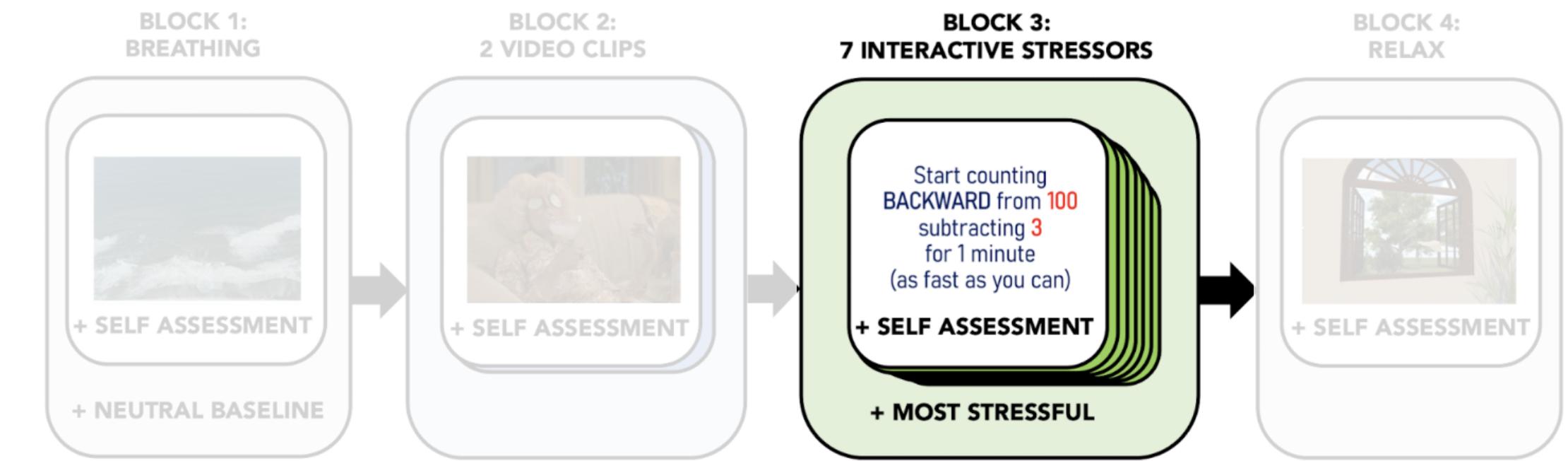
Block 3 : Interactive Stressors

Interactive Stressors

스트레스 유도를 위한 7개의 인터랙티브 스트레스 요인으로 구성
모든 태스크는 1분 내에 응답 해야함
순서는 참가자들이 예상치 못하도록 설계

실험 프로토콜 - 3

Protocol

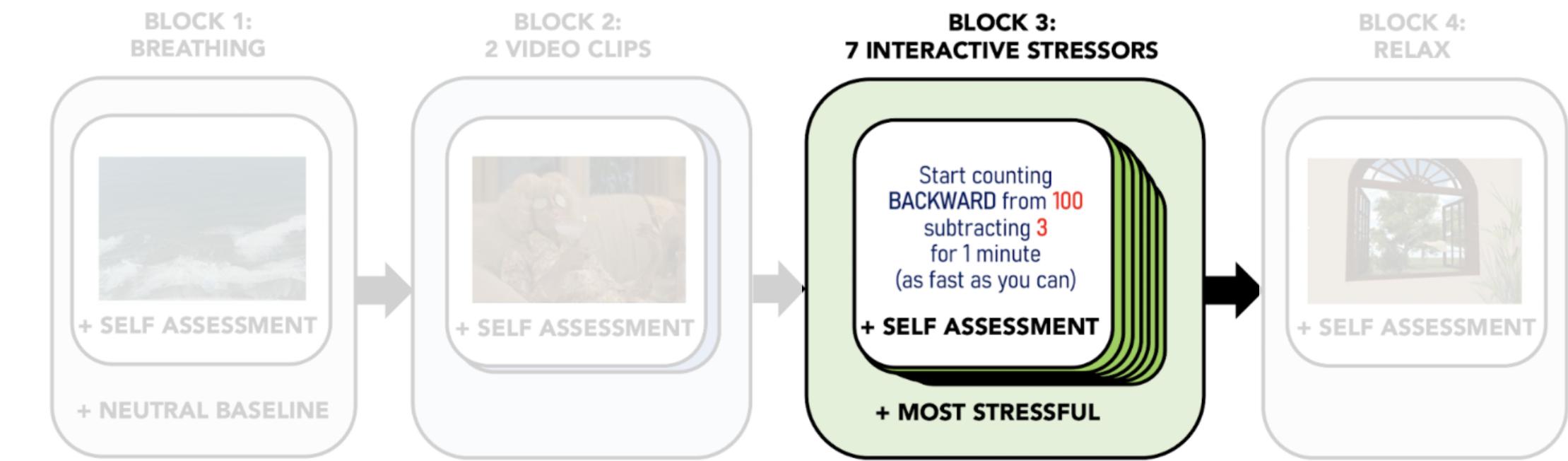


Block 3:
Interactive Stressors



실험 프로토콜 - 3

Protocol



Counting 1

산술 연산을 통해 인지 부하를 증가시키기 위해 설계

참가자들은 100부터 시작해 빠르게 3씩 빼면서 거꾸로 세야함

Counting

Counting 2

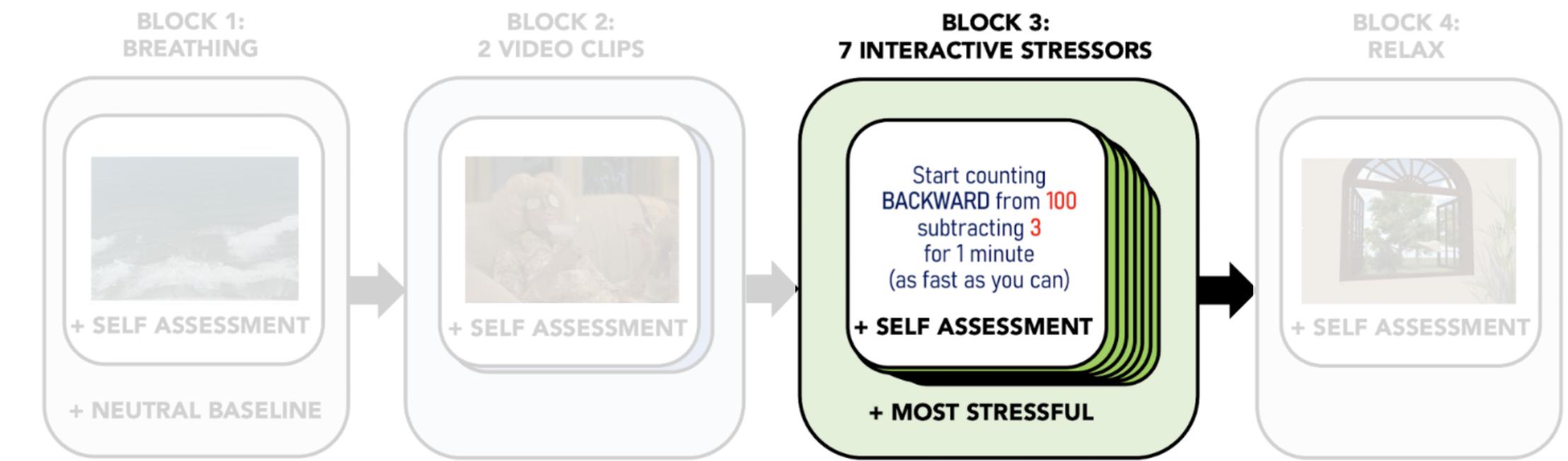
참가자들은 1011부터 시작해 빠르게 7씩 빼면서 거꾸로 세야함

Counting 3

참가자들은 115부터 시작하여 가능한 한 빠르게 3씩 빼면서 거꾸로 셈
동시에 독립적인 손 움직임을 반복

실험 프로토콜 - 3

Protocol

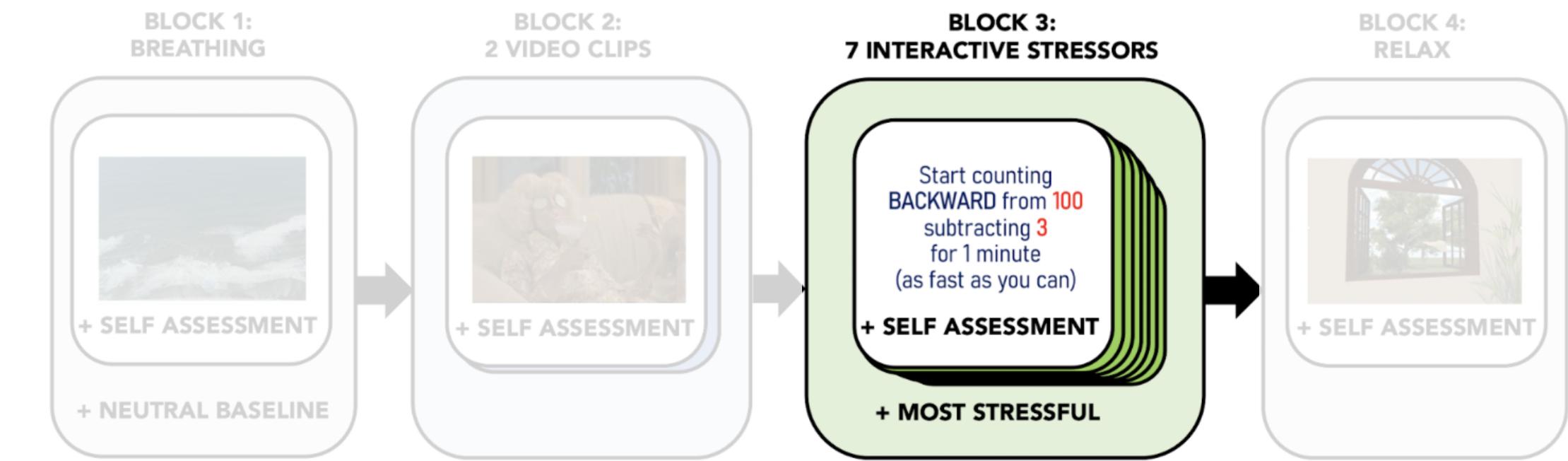


Counting



실험 프로토콜 - 3

Protocol



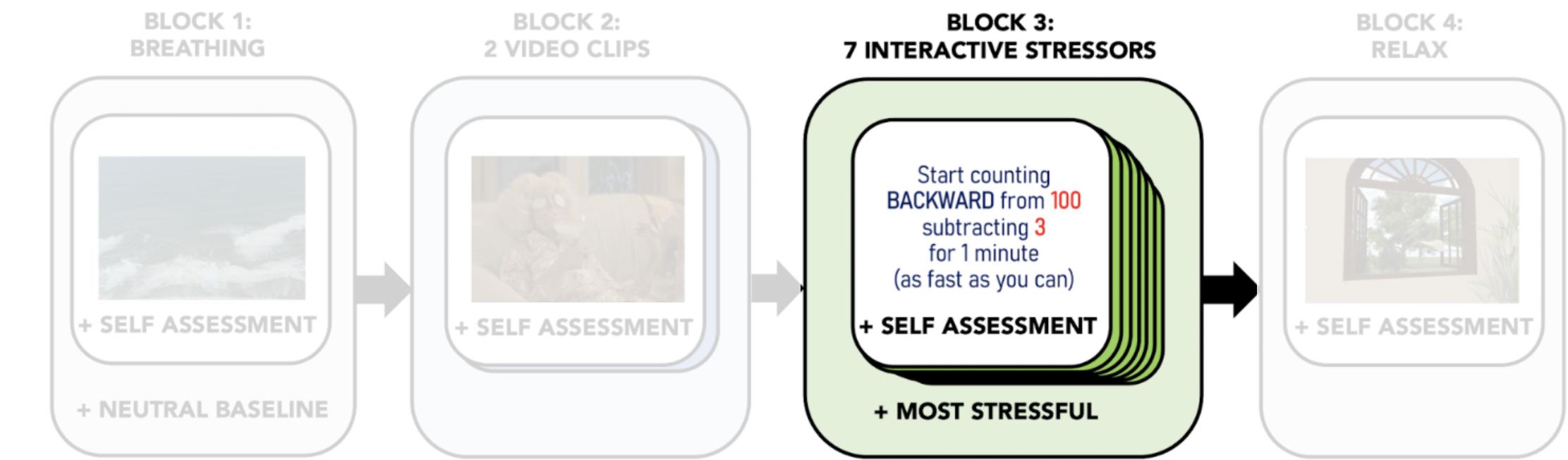
Purple Red Brown
Red Green Blue

Stroop

참가자들의 주의력과 반응성을 요구하여 인지 부하를 증가

실험 프로토콜 - 3

Protocol

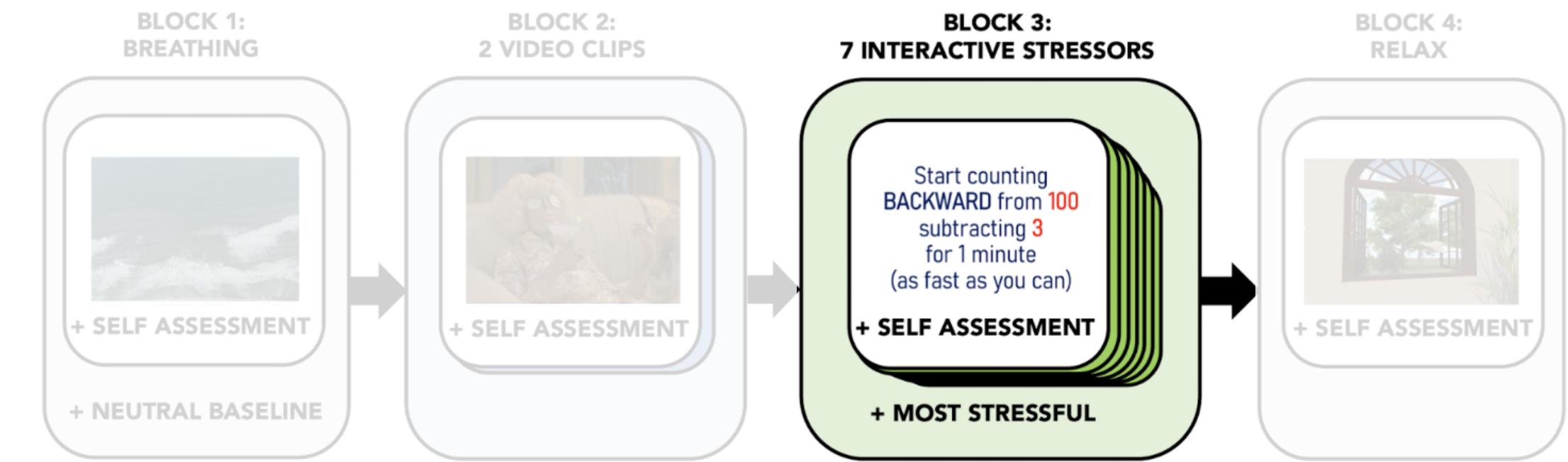


Speaking

자신의 강점과 약점을 설명하도록 지시받아
스트레스가 많은 면접 상황을 모방

실험 프로토콜 - 3

Protocol

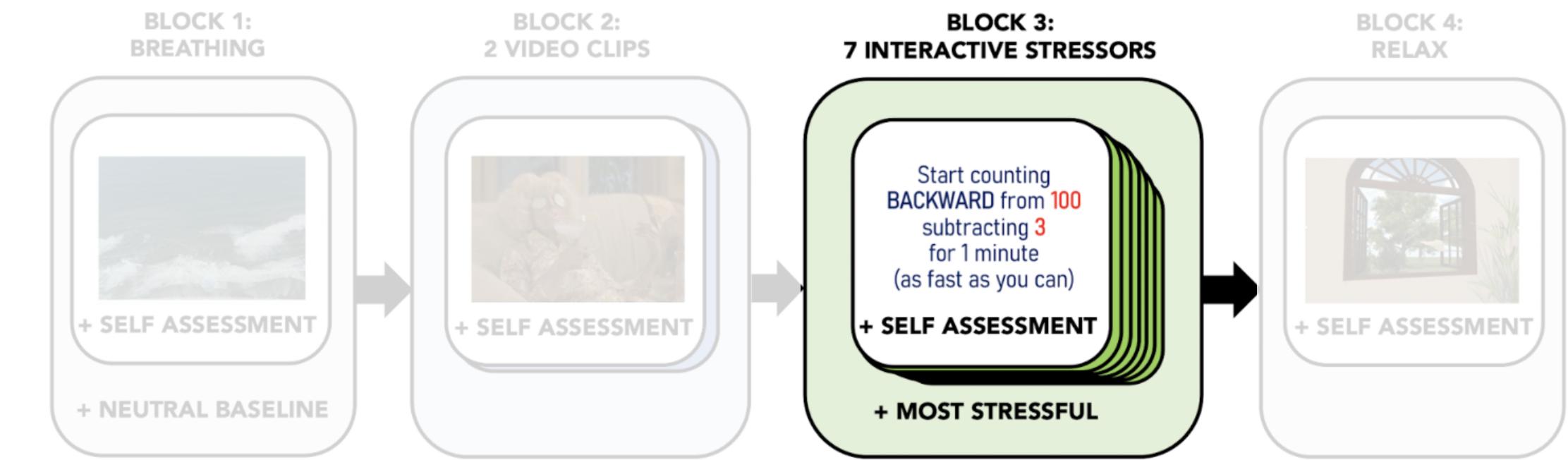


Speaking



실험 프로토콜 - 3

Protocol

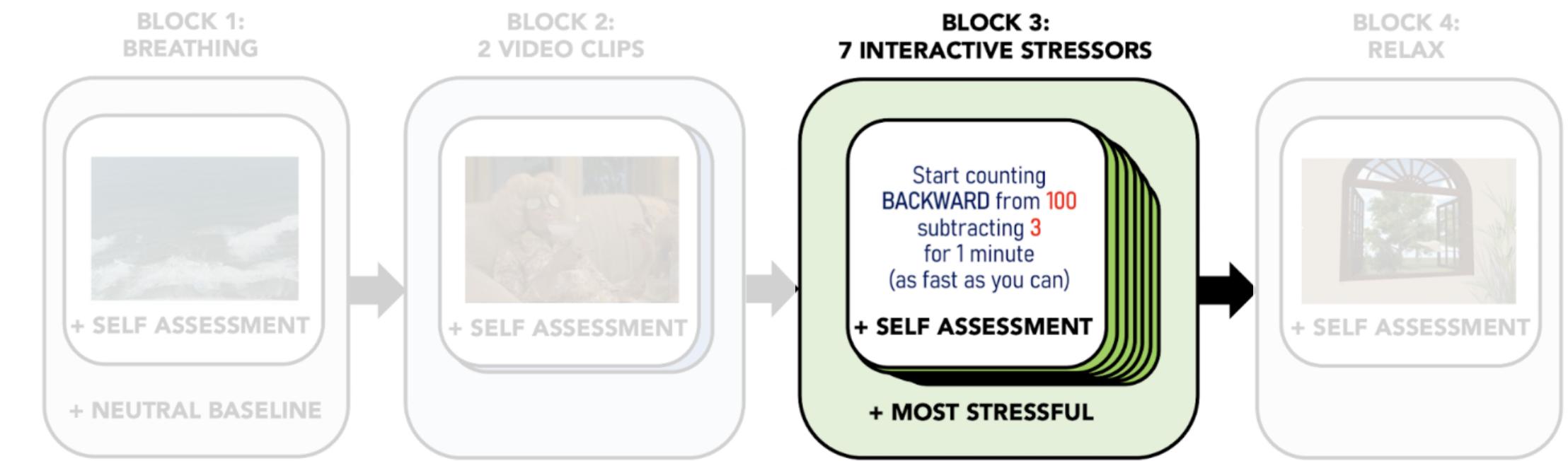


Reading

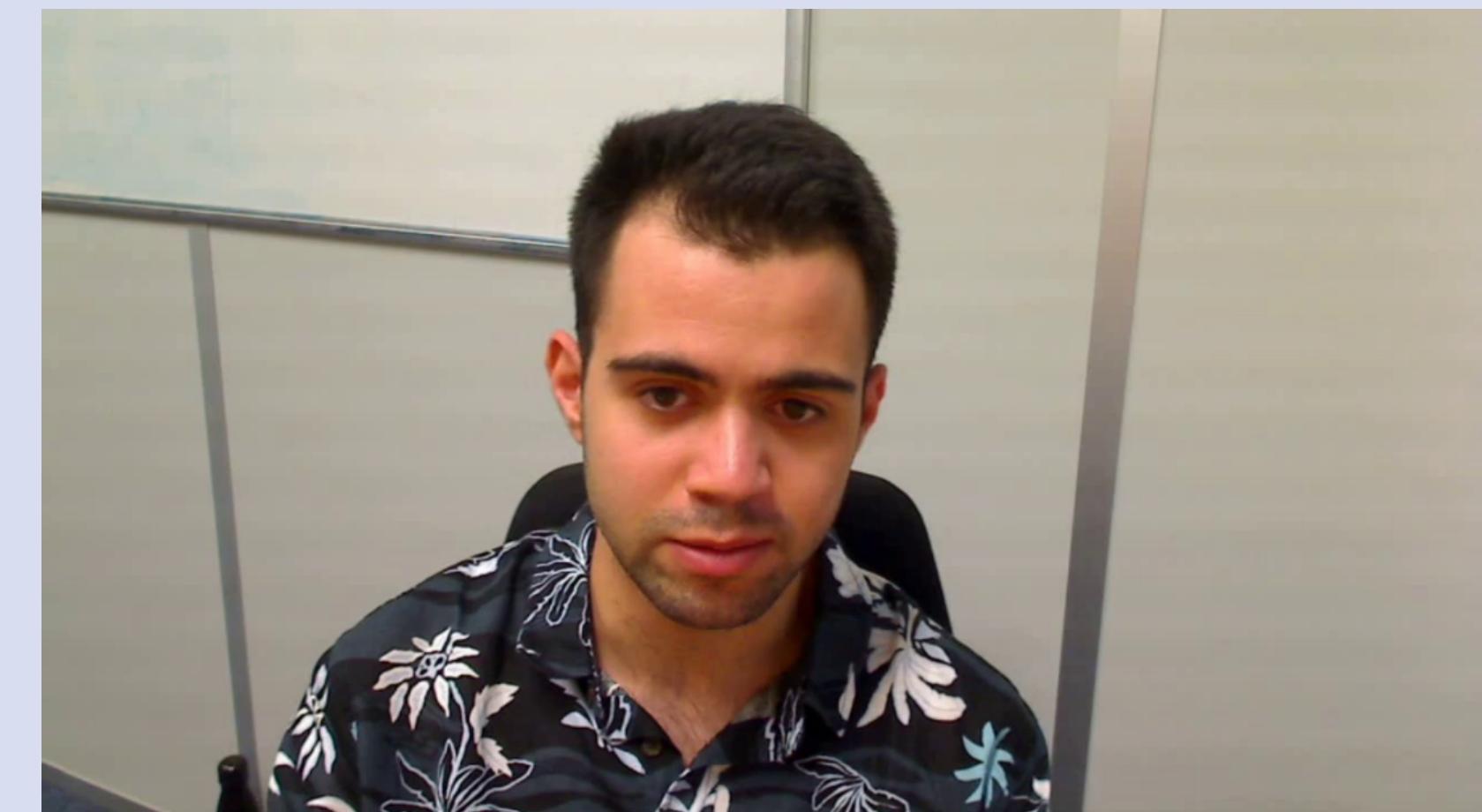
주어진 텍스트를 읽고, 읽은 내용을 설명
이해 능력과 말하기를 동시에 스트레스 요인으로 작용

실험 프로토콜 - 3

Protocol

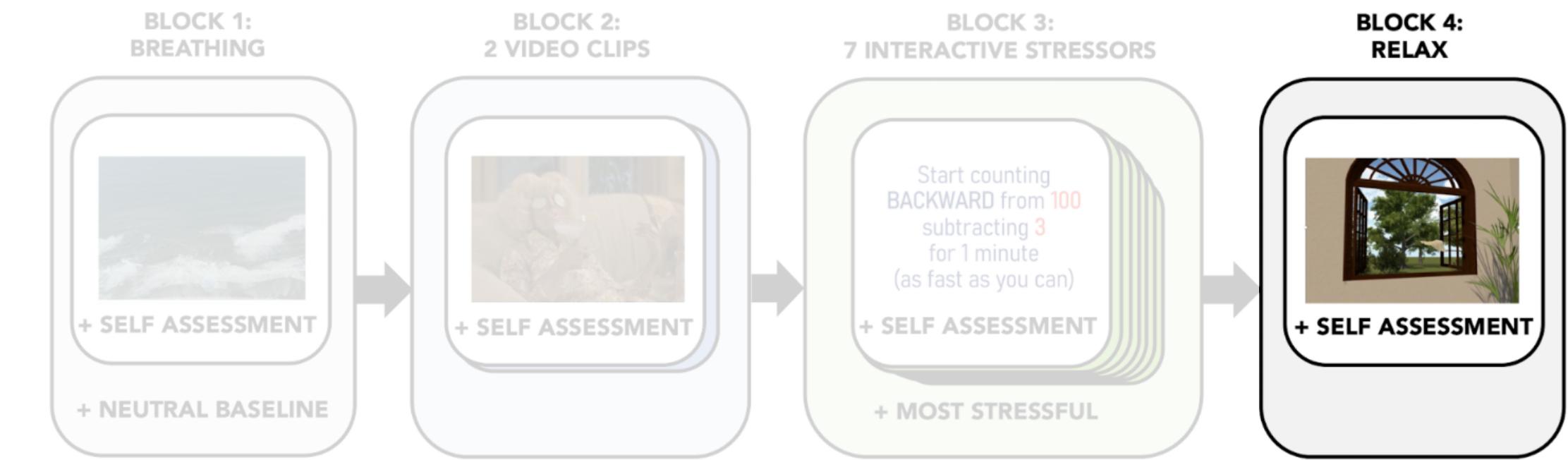


Reading



실험 프로토콜 - 4

Protocol

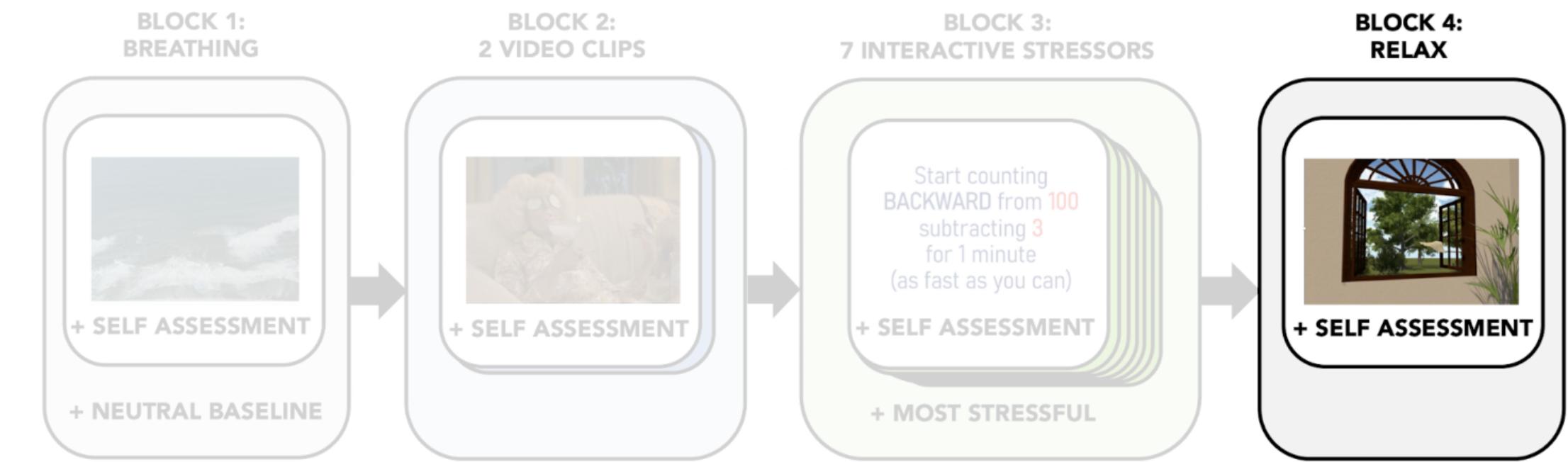


Relax

평화로운 비디오 시청하면서 진정

실험 프로토콜 - 4

Protocol



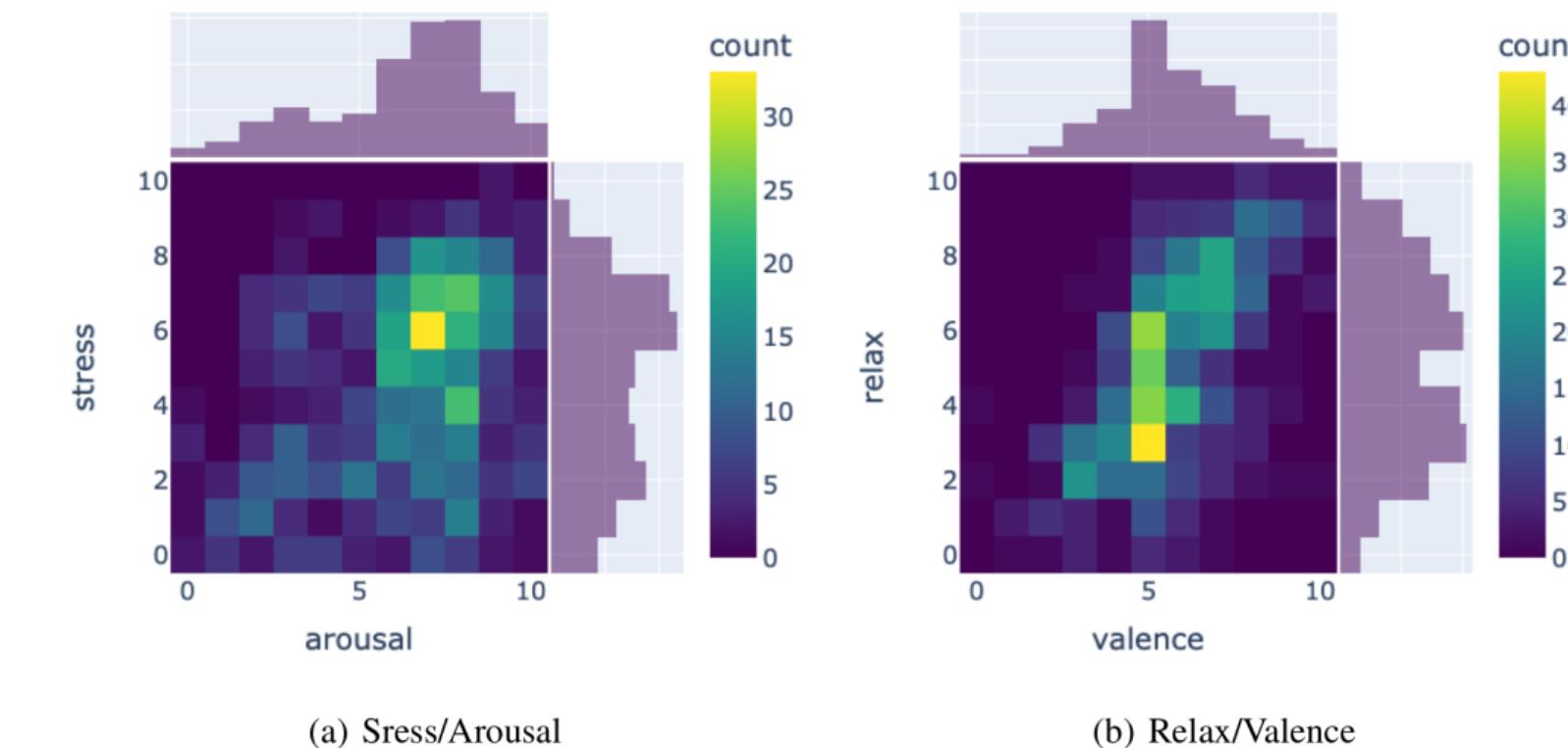
Relax



데이터 라벨링

Data Annotation

Data Annotation



```
get_labels(stress, relaxation, valence, arousal):
```

Data Annotation (self-assessment)

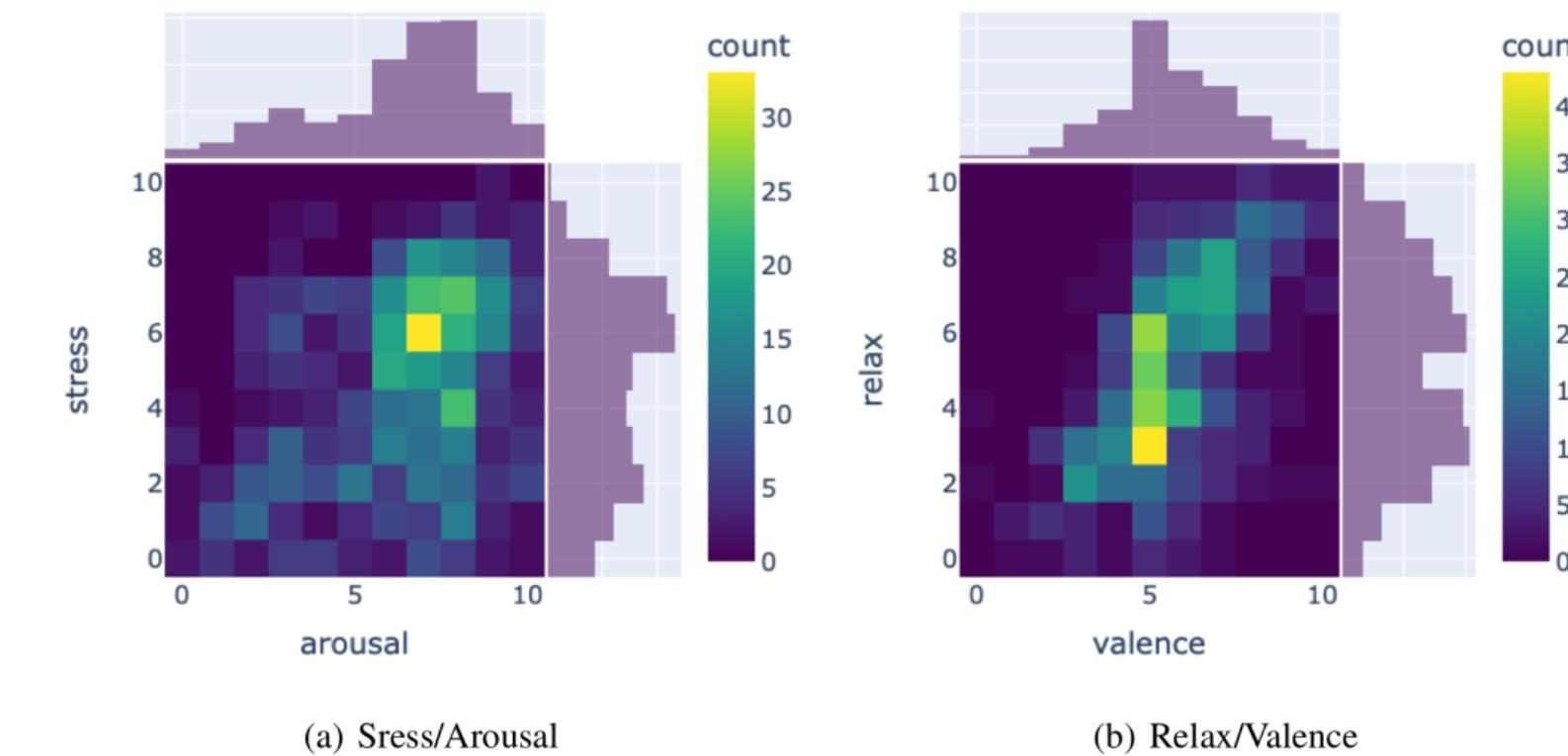
Stress, Relaxation, Valence, Arousal
수준을 각각 0~10점까지 평가

데이터 라벨링

Data Annotation

labels.csv

```
# 2-클래스 라벨
if stress < 5:
    label_2class = 0 # 스트레스 없음
else:
    label_2class = 1 # 스트레스 있음
```

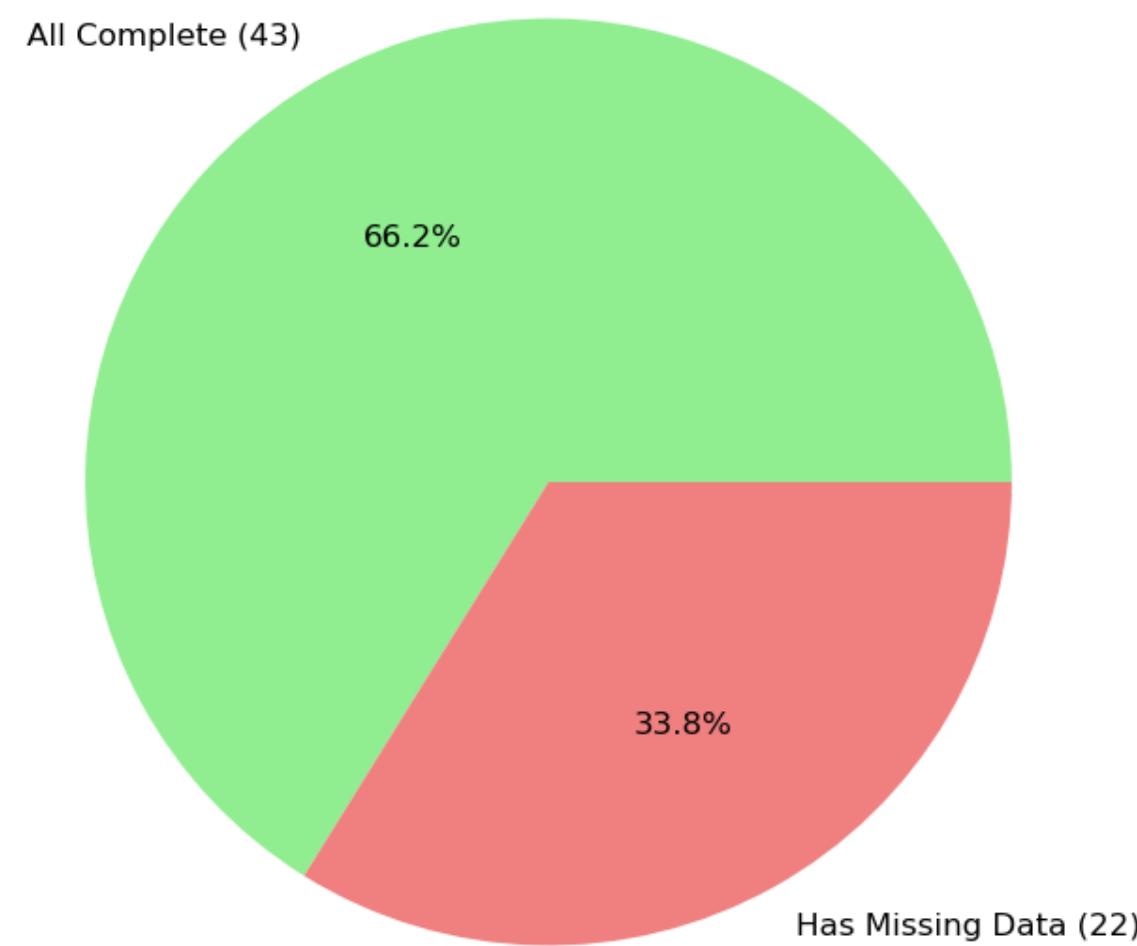


```
# 3-클래스 라벨 (0=안=0, 중립=1, 스트레스=2)
if (valence > 5) and (arousal < 5) and (relaxation > 5):
    label_3class = 0 # 안
elif (arousal > 5) and (valence < 5) and (relaxation < 5):
    label_3class = 2 # 스트레스
else:
    label_3class = 1 # 중립
```

EDA

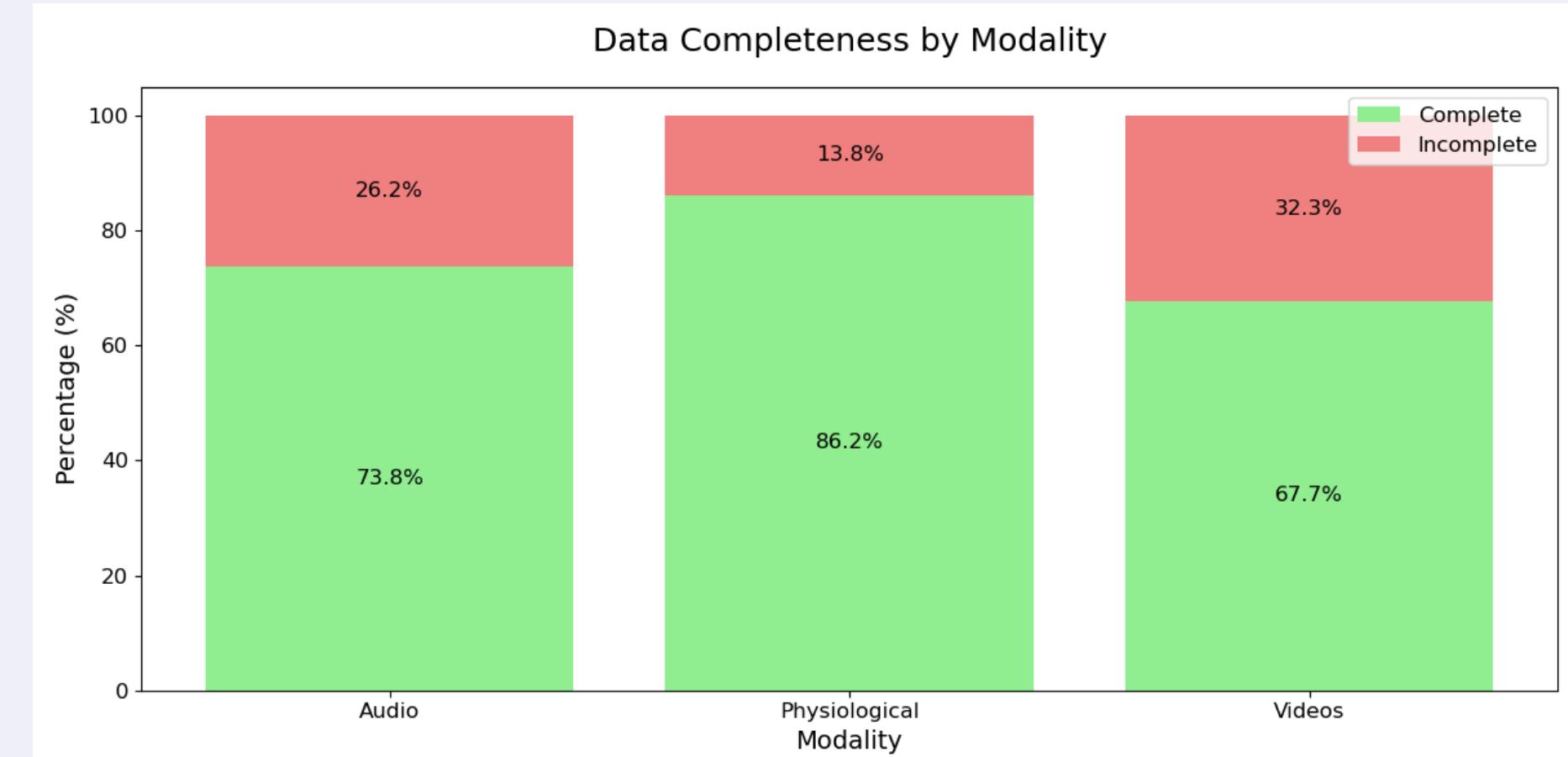
EDA data

Overall Dataset Completeness



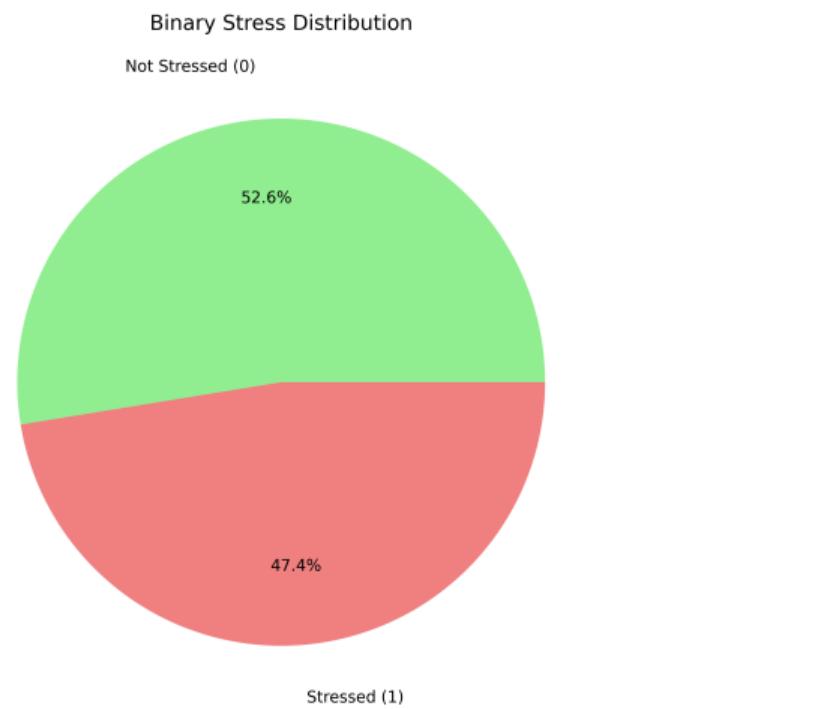
● 완전한 데이터 vs 하나라도 누락된 데이터

총 65명의 subject 중, 43명의 subject가 완전한 데이터

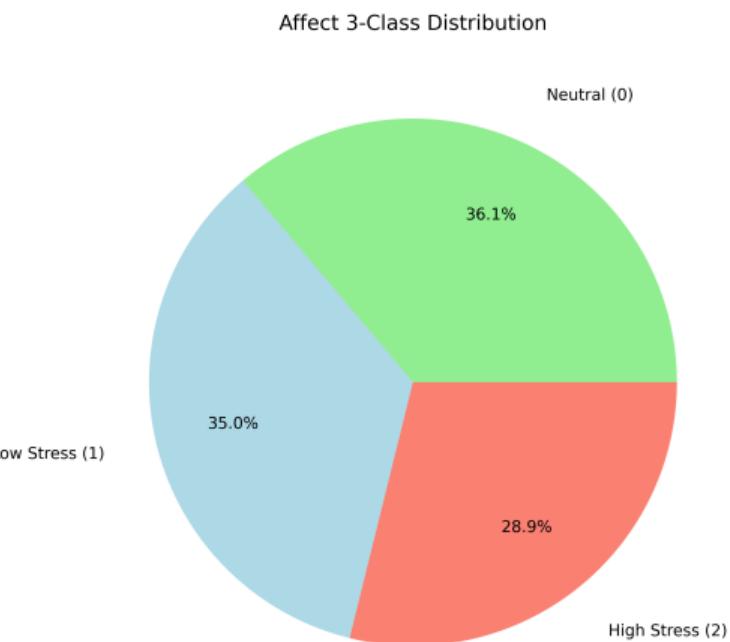


EDA

EDA data



```
# 2-클래스 라벨
if stress < 5:
    label_2class = 0 # 스트레스 없음
else:
    label_2class = 1 # 스트레스 있음
```



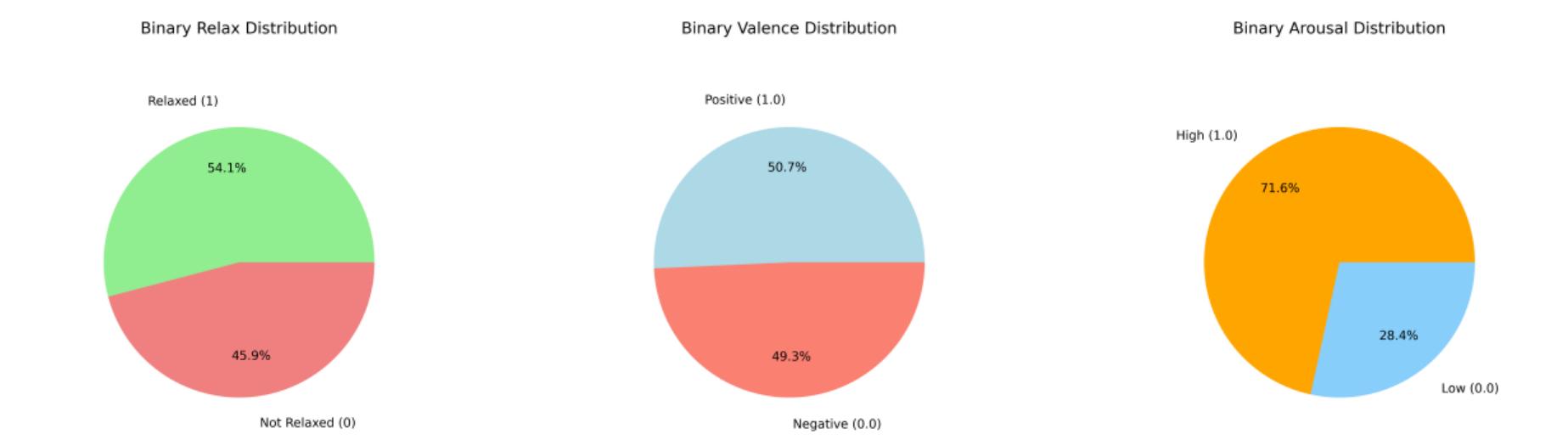
```
# 3-클래스 라벨 (이완=0, 중립=1, 스트레스=2)
if (valence > 5) and (arousal < 5) and (relaxation > 5):
    label_3class = 0 # 이완
elif (arousal > 5) and (valence < 5) and (relaxation < 5):
    label_3class = 2 # 스트레스
else:
    label_3class = 1 # 중립
```

● 라벨링 된 데이터의 비율

참가자들은

Stress, Valence, Arousal, Relaxation 4가지만
0~10점 사이로 자기평가 수행

이것들을 바탕으로 라벨링이 진행됨



연구 방향 제안

Purpose

항목	세부내용	비고
Missing Modalities	<p>Masked Multimodal Transformer Cross-modal Consistency (CLIP-like) 기법 등을 활용하여 각 모달리티 간 상호 보완 및 일관성을 확보하는 방법 제안</p>	<p>Masked Multimodal Transformer 방법은 논문 1개 있음</p>
Multitask	<p>하나의 공통 Encoder를 사용하고 각 태스크 (Stress, Relaxation, Valence, Arousal) 별로 개별 출력 헤드를 구성하는 멀티태스크 아키텍쳐 설계</p>	<p>GOPT에서 영감을 받음</p>

Missing Modalities

Multimodal Learning for Detecting Stress under Missing Modalities

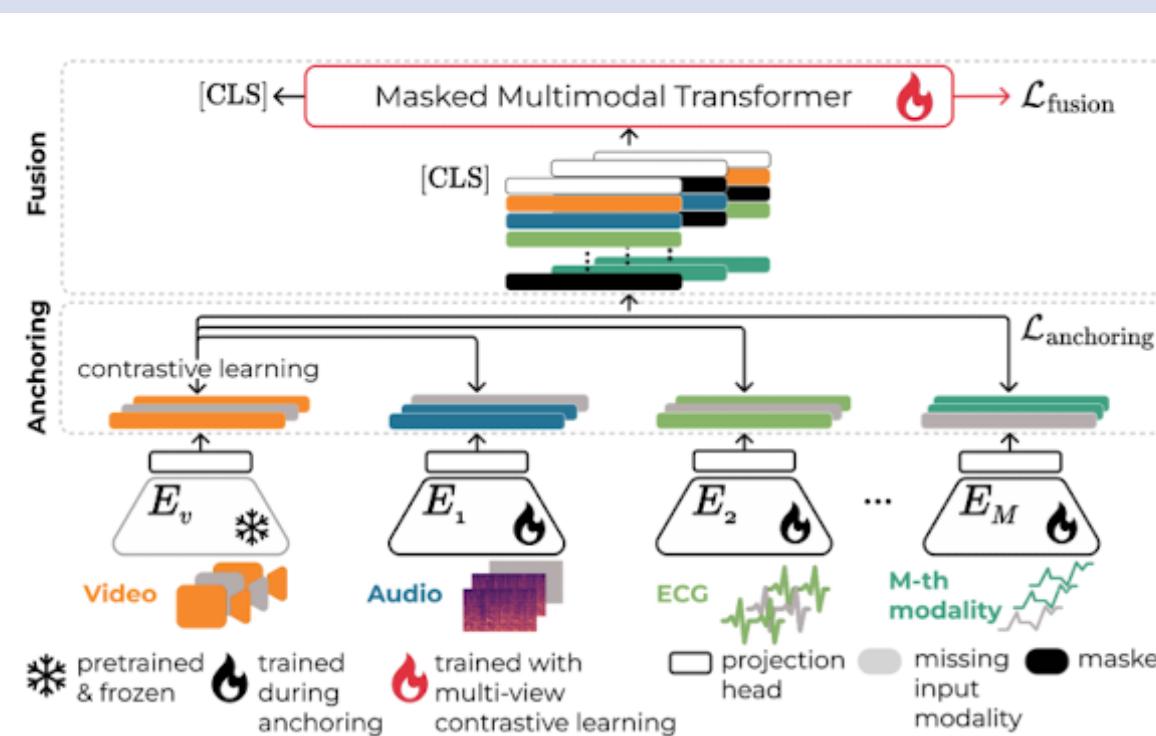


Figure 1. Overview of our method. (i) Contrastive learning aligns unimodal representations to the video. (ii) Fusion models multimodal interactions and handles missing modalities.

Multimodal Learning for Detecting Stress under Missing Modalities

Julie Mordacq^{1,2,4}

Leo Milecki^{1,3}

Maria Vakalopoulou^{1,3}

Steve Oudot^{1,2}

Vicky Kalogeiton^{2,4}

¹Inria Saclay ²Ecole Polytechnique ³CentraleSupélec, Paris-Saclay University ⁴LIX, CNRS, IP Paris

WiCV 2024 - Women in Computer Vision 워크숍 발표 논문

모달리티 정렬 (Anchoring)

가장 강력한 모달리티(논문에서는 비디오를 기준으로 삼음)를 기준 삼아 다른 모달리티의 임베딩을 공통의 임베딩 공간에 정렬
같은 샘플에서 나온 서로 다른 모달리티의 표현 벡터에 일관성 부여

Masked Multimodal Transformer

Transformer 기반의 Fusion 모델을 사용

누락된 모달리티가 있어도 남은 모달리티 간의 상호작용을 통해

스트레스 예측에 필요한 통합 표현 학습

마스킹 기법을 도입해, 누락된 데이터를 어텐션 계산에서 배제

Missing Modalities

Multimodal Learning for Detecting Stress under Missing Modalities

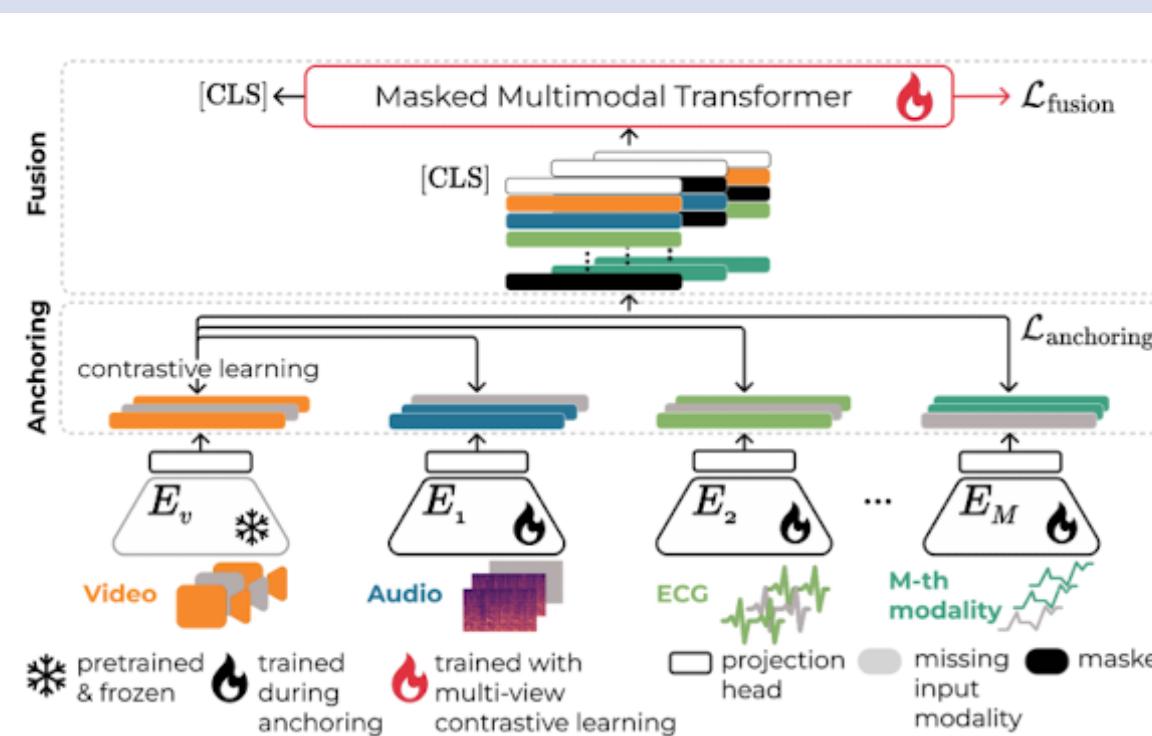


Figure 1. Overview of our method. (i) Contrastive learning aligns unimodal representations to the video. (ii) Fusion models multimodal interactions and handles missing modalities.

Multimodal Learning for Detecting Stress under Missing Modalities

Julie Mordacq^{1,2,4}

Leo Milecki^{1,3}

Maria Vakalopoulou^{1,3}

Steve Oudot^{1,2}

Vicky Kalogeiton^{2,4}

¹Inria Saclay ²Ecole Polytechnique ³CentraleSupélec, Paris-Saclay University ⁴LIX, CNRS, IP Paris

WiCV 2024 - Women in Computer Vision 워크숍 발표 논문

모달리티 정렬 (Anchoring)

가장 강력한 모달리티(논문에서는 비디오를 기준으로 삼음)를 기준 삼아 다른 모달리티의 임베딩을 공통의 임베딩 공간에 정렬
같은 샘플에서 나온 서로 다른 모달리티의 표현 벡터에 일관성 부여

$$\mathcal{L}_{\mathcal{V}, \mathcal{M}_m} = - \sum_{i=1}^B \log \frac{\exp(\cos(f_v^i, f_m^i)/\tau)}{\sum_{k=1}^B \exp(\cos(f_v^i, f_m^k)/\tau)}, \tau \in \mathbb{R}^+$$

Missing Modalities

Multimodal Learning for Detecting Stress under Missing Modalities

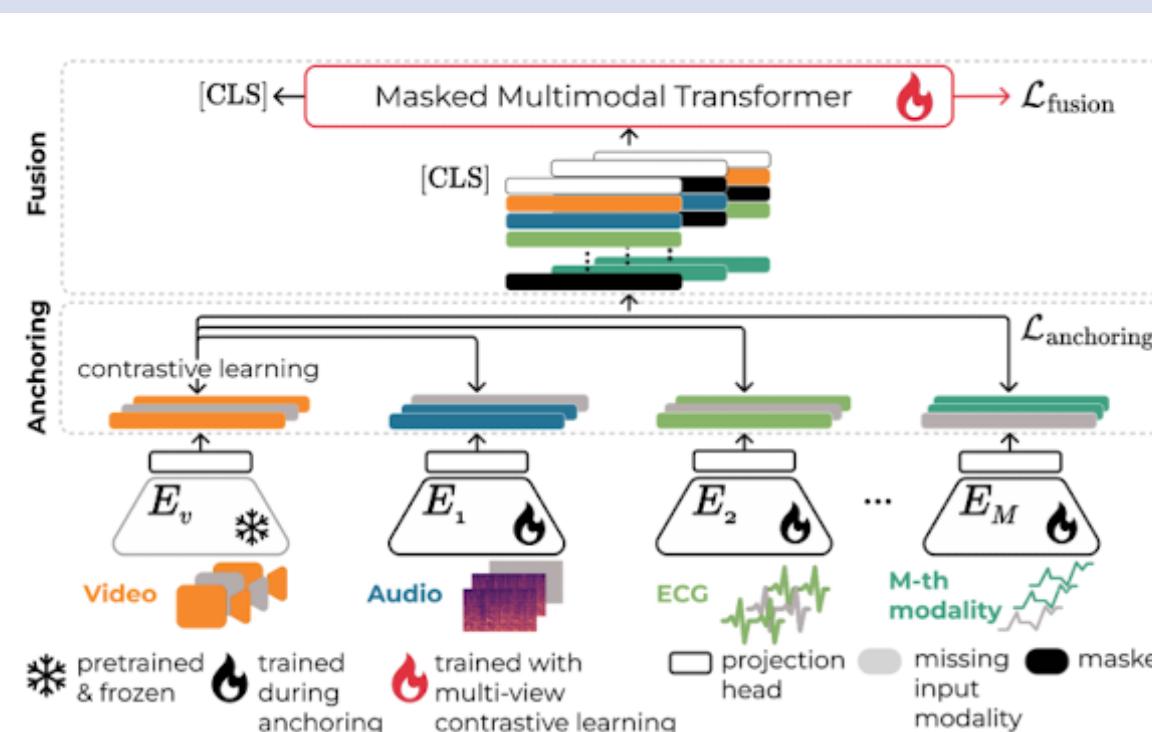


Figure 1. Overview of our method. (i) Contrastive learning aligns unimodal representations to the video. (ii) Fusion models multimodal interactions and handles missing modalities.

Multimodal Learning for Detecting Stress under Missing Modalities

Julie Mordacq^{1,2,4}

Leo Milecki^{1,3}

Maria Vakalopoulou^{1,3}

Steve Oudot^{1,2}

Vicky Kalogeiton^{2,4}

¹Inria Saclay ²Ecole Polytechnique ³CentraleSupélec, Paris-Saclay University ⁴LIX, CNRS, IP Paris

WiCV 2024 - Women in Computer Vision 워크숍 발표 논문

Masked Multimodal Transformer

Transformer 기반의 Fusion 모델을 사용

누락된 모달리티가 있어도 남은 모달리티 간의 상호작용을 통해
스트레스 예측에 필요한 통합 표현 학습

마스킹 기법을 도입해, 누락된 데이터를 어텐션 계산에서 배제

$$O_i = \sum_j z_{ij} \frac{\exp(Q_i^T K_j / \sqrt{d_k})}{\sum_{\{j', z_{ij'}=1\}} \exp(Q_i^T K_{j'} / \sqrt{d_k})} V_j$$

Missing Modalities

Multimodal Learning for Detecting Stress under Missing Modalities

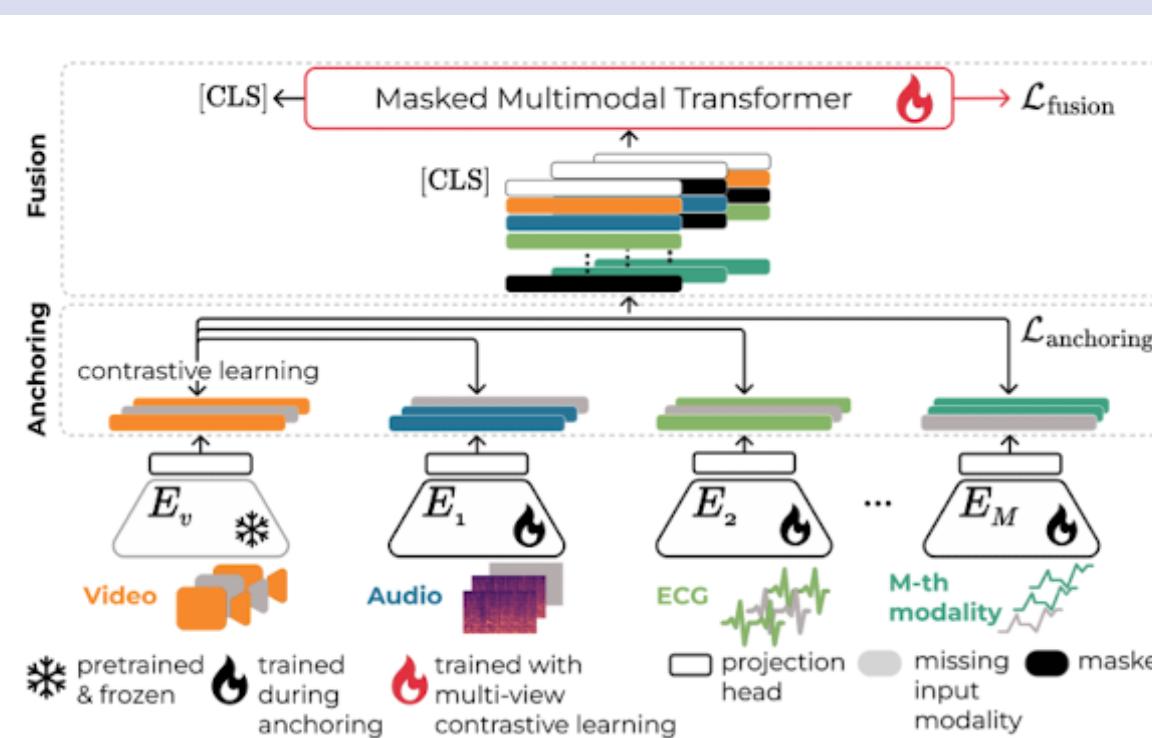


Figure 1. Overview of our method. (i) Contrastive learning aligns unimodal representations to the video. (ii) Fusion models multi-modal interactions and handles missing modalities.

Multimodal Learning for Detecting Stress under Missing Modalities

Julie Mordacq^{1,2,4}

Leo Milecki^{1,3}

Maria Vakalopoulou^{1,3}

Steve Oudot^{1,2}

Vicky Kalogeiton^{2,4}

¹Inria Saclay ²Ecole Polytechnique ³CentraleSupélec, Paris-Saclay University ⁴LIX, CNRS, IP Paris

WiCV 2024 - Women in Computer Vision 워크숍 발표 논문

	ACC	F1
Video	62(4) [‡]	67(3) [‡]
Biomedical signals	58(4) [‡]	66(5) [‡]
Audio	62(4) [‡]	67(4) [‡]
Feature Fusion [1]	61(3) [‡]	66(4) [‡]
Decision Fusion [†] [1]	65(5) [‡]	72(5) [‡]
<i>Ours</i>	69.5(3.7)	75.9(4.3)

Missing Modalities

Multimodal Learning for Detecting Stress under Missing Modalities

Multimodal Learning for Detecting Stress under Missing Modalities

Julie Mordacq^{1,2,4}

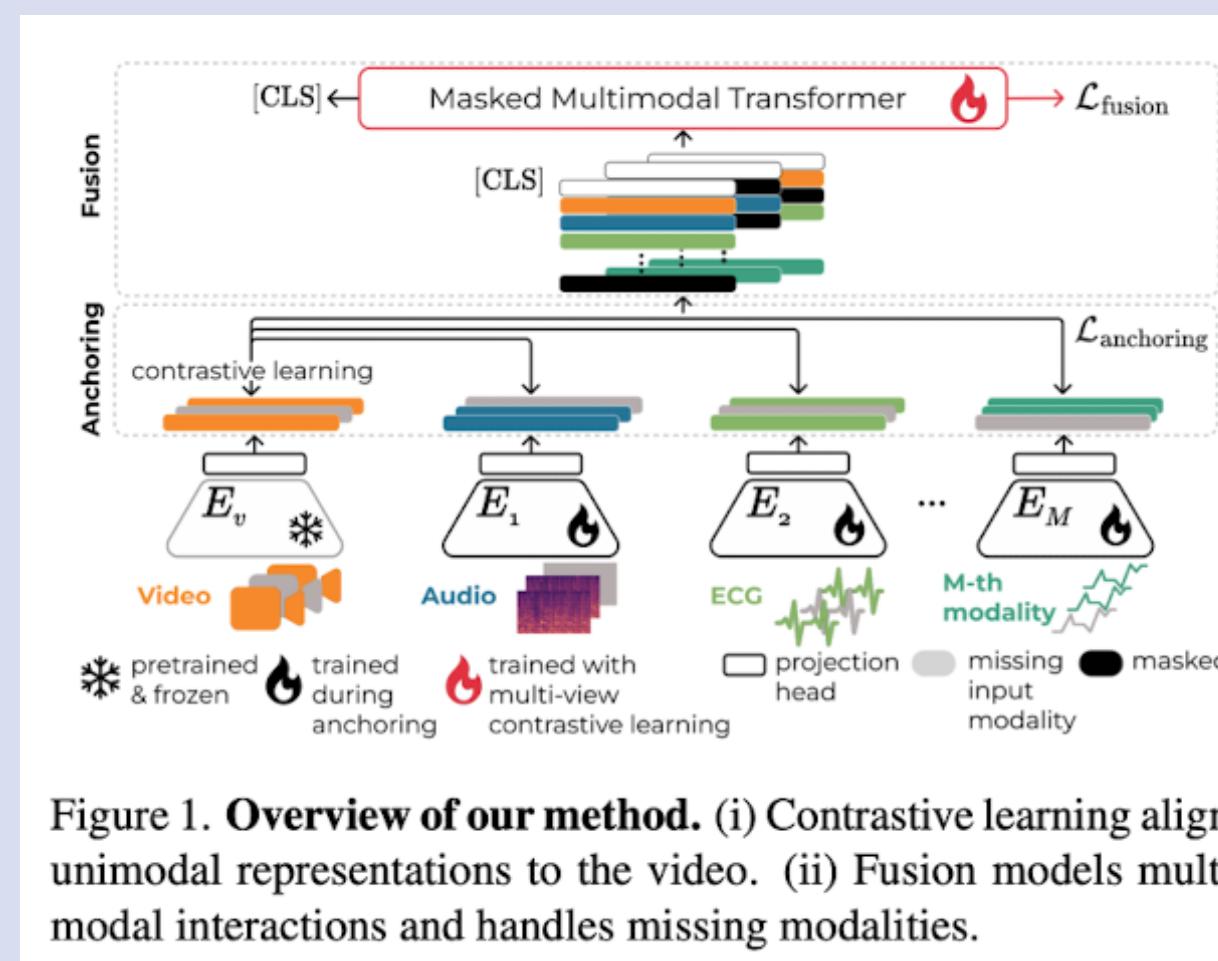
Leo Milecki^{1,3}

Maria Vakalopoulou^{1,3}

Steve Oudot^{1,2}

Vicky Kalogeiton^{2,4}

¹Inria Saclay ²Ecole Polytechnique ³CentraleSupélec, Paris-Saclay University ⁴LIX, CNRS, IP Paris



WiCV 2024 - Women in Computer Vision 워크숍 발표 논문

논문에서 제시한 누락 모달리티 처리 방법인
Masked Multimodal Transformer 대신,
Cross-modal Consistency (CLIP-like) 을 적용하여
서로 다른 모달리티의 임베딩을 공통 공간으로 정렬하는 방식도
재미있을 것으로 생각

Multitask

TRANSFORMER-BASED MULTI-ASPECT MULTI-GRANULARITY NON-NATIVE ENGLISH SPEAKER PRONUNCIATION ASSESSMENT

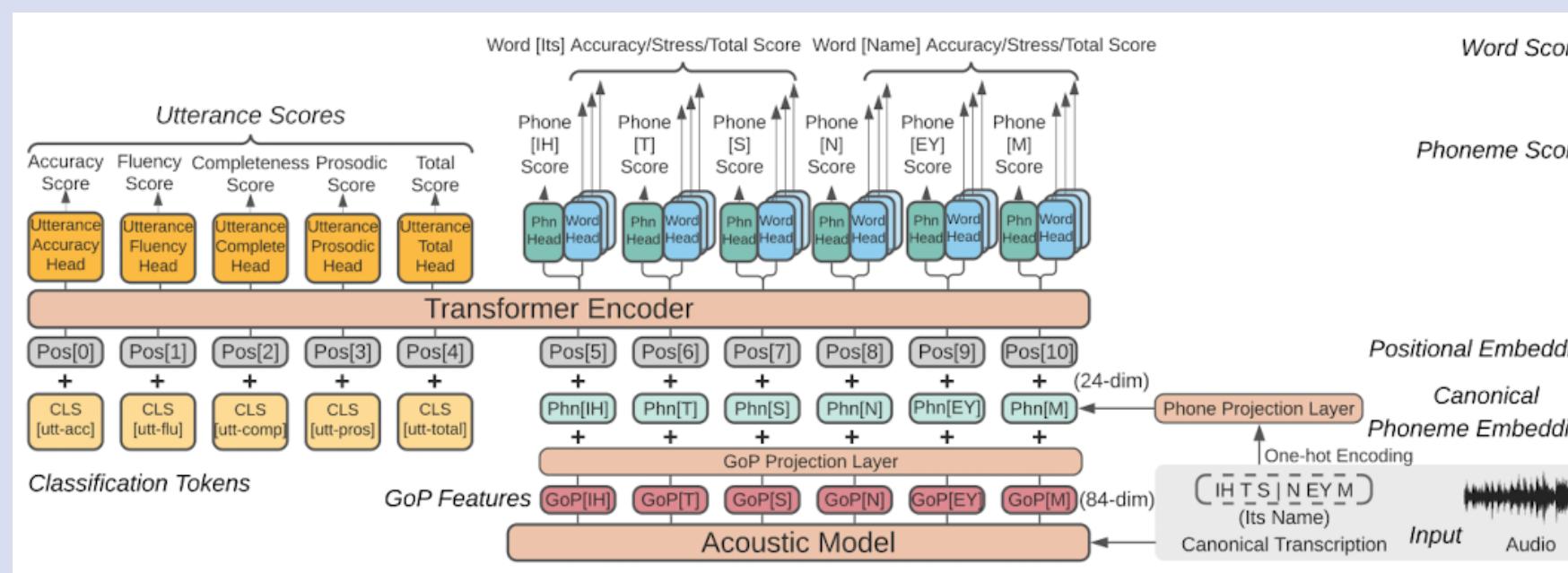
TRANSFORMER-BASED MULTI-ASPECT MULTI-GRANULARITY NON-NATIVE ENGLISH SPEAKER PRONUNCIATION ASSESSMENT

Yuan Gong¹, Ziyi Chen², Iek-Heng Chu², Peng Chang², James Glass¹

¹MIT CSAIL, Cambridge, MA 02139, USA

²PAII Inc., Palo Alto, CA 94306, USA

{yuangong,glass}@mit.edu {chenziyi253,zhuyixing276,changpeng805}@pingan.com.cn



현장 실습 때 참고했던 GOPT 논문

01 GOPT에 유사하게

4개의 [CLS] 토큰을 두어 Stress, Valence, Arousal, Relaxation
동시에 예측하는 멀티태스크 모델을 개발

StressID + Multitask 연구는 아예 없음

Thank You

김동환