

분류		Data 이름	설명	Publication	비고
Raw Data / Behavior, Brain Function, Brain Structure, Connectivity	1	7T audio-only movie fMRI (+cardiac/resp)	Audio-visual보다 먼저 촬영됨. 정확히 같은 Movie cut을 사용. (N=20)	Hanke et al.(2014), A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie, <i>Scientific Data</i>	
	2	7T Listening to music task fMRI (+cardiac/resp)	25 music clips with/without speech	Hanke et al.(2015), High-resolution 7-Tesla fMRI data on the perception of musical genres, <i>F1000Research</i> (+ Sengupta, Pollmann & Hanke(2018), Spatial band-pass filtering aids decoding musical genres from auditory cortex 7T fMRI, <i>F1000Research</i> ) ??	
	3	3T Retinotopic mapping task fMRI [phase2]	Eccentricity & polar angle maps of the V1	Sengupta et al.[Hanke](2016), A studyforrest extension, retinotopic mapping and localization of higher visual areas, <i>Scientific Data</i> (preprint: Sengupta et al.(2016), An extension of the studyforrest dataset for vision research, <i>bioRxiv</i> )	
		3T Higher visual area localizer task fMRI [phase2]	얼굴, 머리 없는 몸, 작은 사물, 바깥 장면		
	4	3T audio-visual movie fMRI, eyegaze (+cardiac/resp) [phase2]	T2*-weighted EPI, TR=2s, TE=30s, 35 slices with 80x80 voxels, 3mm each 451, 441, 438, 488, 462, 439, 542, 338 volumes per segments 1-8 respectively / 약 54 000 time points Audio-only를 촬영한 기준 15명에 대하여 추가로 촬영한 버전.	Hanke et al.(2016), A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation, <i>Scientific Data</i> (preprint: Hanke et al.(2016), Simultaneous fMRI and eye gaze recordings during prolonged natural stimulation - a studyforrest extension, <i>bioRxiv</i> ) [github] studyforrest-data-phase2 → An extension of studyforrest.org data set → simultaneous fMRI/eyetracking while movie watching, plus visual localizers [OpenFMRI] ds000113d / [OpenNeuro] ds000013 (원본, b, c, d 통합버전)	sub-<ID>/ses-movie 디렉토리 → *ses-movie_task-movie*_bold.nii.gz ds00013(d)
	5	3T/7T Visual orientation task fMRI (= multires7T)	몇 명 없음	[7T] Sengupta et al.[Hanke](2017), Ultra high-field(7T) multi-resolution fMRI data for orientation decoding in visual cortex, <i>Data in Brief</i> [7T] (+ Sengupta et al.(2017), The effect of acquisition resolution on orientation decoding from v1 bold fMRI at 7T, <i>NeuroImage</i> ) ?? [3T] (+ Sengupta et al.(2018), The effect of acquisition resolution on orientation decoding from V1: comparison of 3T and 7T, <i>bioRxiv</i> )	
Pre-processed Data	6	T1, T2, DWI, SWI, Angiography		-	
	1	Audio-visual movie fMRI aggregate ROI timeseries	For Shen et al.(2013) cortex parcellation 3T로 찍은 phase2 데이터인 듯! (3Tp2)	-	
	2	Per-participant aligned fMRI data	for within-subject analysis		
	+	Denoised 3T audio-visual movie fMRI dataset (N=15, 2h, 4번을 그대로 사용, 각 run마다 denoise)	각 참가자의 각 run에 대하여 (1)전처리, (2)Spatial Independent Component Analysis를 통한 decomposition, (3)신호/노이즈의 manual classification, (4)artifact removal을 거친 데이터. 기존 데이터에 비해 시간 SNR이 높고 ICA(Inter-subject Correlation Analysis)의 감도도 좋음	Liu et al.(2019), A manually denoised audio-visual movie watching fMRI dataset for the studyforrest project, <i>Scientific Data</i> Denoised fMRI data Location: sub-*_ses-movie/fnc/sub-*_ses-movie_task-movie_run-<run_id>_space-T1w_desc-preproc_desc-[sm5,unsm]_desc-denoised_bold.nii.gz	phase2-denoised-openneuro repo ds001769
	3	Participant and scan-specific template MRI images	For alignment, masking, structural properties & (non-)linear transformations between image spaces		
Movie Stimulus Annotations	4	Freesurfer cortical surface reconstruction		-	
	1	Cuts, depicted locations, temporal progression	870개 shot의 정확한 타이밍, 장소에 대한 설명(주상화 정도 고-중-저, 총 3개의 version), 0이야기 shot 사이의 시간 차이를 묘사하는 4개의 classes. 각 shot은 안/밖, 하루 중 시간으로 label됨. Stimulus timing은 BIDS' *_events.tsv 파일들을 참고(6-column text files, including onset, duration, frame index[frameidx] of each movie frame)	Hausler & Hanke(2016), An annotation of cuts, depicted locations, and temporal progression in the motion picture "Forrest Gump", <i>F1000Research</i>	
	2	Speech	2,500개 spoken sentences 16,000개 words (202개의 non-speech vocalization 포함) 66,000개 phonemes들의 정확한 타이밍 + 발화자 + 각 단어의 품사(명사, 동사, 형용사 등) + 문장성분 관계{주어, 동사, 목적어 등}가 정의됨.	Hausler & Hanke(2021), A studyforrest extension, an annotation of spoken language in the German dubbed movie "Forrest Gump" and its audio-description, <i>F1000Research</i>	
	3	Portrayed emotions	영화와 audio description stimulus에서 묘사된 감정들의 onset, duration, arousal, valence와 직접적인 감정 카테고리 label, 그 감정이 존재한다는 perceptual evidence의 기록	Labs et al.[Hanke](2015), Portrayed emotions in the movie "Forrest Gump", <i>F1000Research</i>	
	+	Perceived emotions	12명의 Italian 참가자들이 영화를 보면서 기록한 continuous emotion ratings(intensity[0~100] for six basic emotions[happiness, surprise, fear, sadness, anger, disgust])	Lettieri et al.(2019), Emotionotopy in the human right temporo-parietal cortex, <i>Nature Communications</i>	직접 다운?
	4	Semantic conflict	거짓말, 아이러니, 비꼬기(sarcasm)를 포함하는 일화들 / 3명의 독립적 관찰자들이 정의	Hanke & Ibe(2016), Lies, irony, and contradiction - an annotation of semantic conflict in the movie "Forrest Gump", <i>F1000Research</i>	
	5	Body contact	body contact 사건들의 타이밍, 행위자, 당하는 사람, 관여하는 신체 부위, 터치의 강도와 valence, 가능성 있는 audio cues.	- (github repo에 있는 todosandnotes 문서를 보니 데이터품질이 의심됨...)	
	6	Eye movement labels	15명은 MRI 스캐너 안에서, 나머지 새로 참가한 15명은 실험실 세팅에서 fMRI 없이 눈 움직임을 기록('누워서 작은 fMRI 화면을 통해 영화를 보는' scanner setting의 영향을 의식하여, 추가로 수집한 control group). Saccades, post-saccadic oscillations, fixations, smooth pursuit event를 분리하여 기록함 / 영화가 complex multi-sensory input을 포함하기 때문에, 참여자가 어디에 집중하고 있는지(attention)를 아는 것이 중요하다는 rationale. / 기기는 Eyelink 1000 사용 / 참가자 ID는 20보다 큼	[Raw Data 4번 (3T audio-visual과 동일)] Hanke et al.(2016), A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation, <i>Scientific Data</i> [Eye detection procedure] Asim et al.(2019), REMoNaV: Robust Eye Movement Detection for Natural Viewing, <i>bioRxiv</i> (preprint)	col 설명은 readme 참조
	7	music	타이밍, 아티스트, 곡 제목, 곡이 발표된 년도		
	8	Low-level perceptual confounds	frame-wise(40밀리초)로 청각/시각적 low-level confound를 기록. ex. 볼륨, 좌-우 볼륨 차이, 각 frame의 밝기, 각 frame의 이전 frame과의 perceptual difference	-	

Reference: [3T fMRI quality metrics] Hanke, M., Adelhöfer, N., Kottke, D. et al. A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Sci Data* 3, 160092 (2016).

## Studyforrest Data Structure

- **.datalad**
  - .gitattributes
  - config

- **artifact**: 모든 획득 데이터에 대한 원시 데이터. Not publicly accessible
  - 3T\_movie\_eyetracking
  - 3T\_structural\_mri
  - 3T\_visuallocalizer
  - 7T\_audiomovie
  - 7T\_musicperception
  - media
  - movie\_eyetracking

- **code / conversion\_qa**: 데이터 구조화, 테스트, 비식별화 변환, 분석에 쓰인 코드.  
쓸 땐 소스 배포의 COPYING 파일을 참조
  - .datalad
  - 3rd: 직접 개발하지 않은 외부 라이브러리나 서드파티 도구인 것으로 추정됨.
  - data / deface: 비식별화 절차에 사용되는 템플릿과 마스크 이미지
  - doc: API 문서의 소스(원본) 파일 (Application Programming Interface Documentation): 개발자가 모듈, 라이브러리 등 코드 기능을 이해하고 올바르게 사용할 수 있도록 도와주는 문서.rst)
  - **gumpdata**: Python 데이터 접근 모듈 (`_init_.py`, `utils.py`, `pymvpa.py` 등)
    - ◆ **tests**: 데이터 무결성, 일관성 검증을 위한 테스트. 데이터 파일이 들은 루트 디렉토리로 이동하여 다음 명령을 실행하면 가능한 모든 테스트를 실행함: `$ PYTHONPATH=<path-to-repo> checkout>:$PYTHONPATH nosetests`
  - **scripts**
    - ◆ **conversion**: 데이터 원시→공개 변환 스크립트
    - ◆ **misc**
    - ◆ **openfmri**: 그룹 템플릿 이미지를 생성하기 위한 `openmri_helpers` 패키지 설정 파일
    - ◆ **parcellation\_ts**
    - ◆ **qa**
    - ◆ **similarity\_analysis**: univariate/multivariate pattern similarity analysis 수행 스크립트
    - ◆ **figures**: 논문의 도표를 만들 때 쓴 가리고 하던데... 없음ㅠㅠ
  - **.gitignore**: git 버전 관리 시스템에서 특정 파일/디렉토리를 무시하도록 설정함.
  - **.gitmodules**: 프로젝트 내에서 다른 git 저장소를 하위 모듈로 포함하도록 설정함.
  - **.noannex**: Datalad와 관련된 파일. 이 디렉토리는 annex가 관리하지 않도록 함.
  - **COPYING**
  - **Makefile**
  - **NOTE**
  - **README.rst**
  - **setup.py**

파란색: **special remote 접근제한/변경/삭제 등으로 다운로드 불가**

- **derivative**: 전처리 데이터, 분석 결과
  - **aggregate\_fmri\_timeseries**: audio-visual movie fMRI aggregate ROI timeseries로 추정됨. Shen et al.(2013) cortex parcellation을 위해 사용됨.
    - ◆ .datalad
    - ◆ atlases / shen: Shen Atlas(ROI: 268개)를 기반으로 한 functional connectivity data (fconn: 두 뇌 영역 간의 상관관계/동시활성패턴/연결강도. 상관계수 행렬로 표현)
      - fconn\_atlas\_150\_1mm.nii.gz
      - fconn\_atlas\_150\_2mm.nii.gz
    - ◆ code (.sh 확장자 bash script가 있음)
    - ◆ src
    - ◆ sub-01 … sub-20
      - atlases / bold3Tp2: 3T MRI phase 2를 의미하는 듯
        - shen\_fconn\_atlas\_150.nii.gz
      - shen\_fconn
        - sub-01\_task-avmovie\_run-1\_bold.csv … /row: timepoint, col: ROI voxel
    - ◆ README.md
  - **aligned\_mri**: '각 참가자에 맞춘' 템플릿 이미지에 정렬된 BOLD fMRI timeseries. templatetransforms 디렉토리에서 제공하는 변환을 사용하여 정렬됨.
    - ◆ code: 모든 파일을 계산하는 데 사용된 소스 코드 및 몇 가지 검증 분석을 포함.
    - ◆ src
    - ◆ sub-01 … 20 / i\_bold3Tp2: 참가자별로 aomovie, avmovie, lclizer, obctgris가 정리.
      - sub-01\_task-aomovie\_run-1\_bold.nii.gz ~ run 8까지
      - sub-01\_task-avmovie\_run-1\_bold.nii.gz & bold\_mcparams.txt ~ run 8까지
      - sub-01\_task-movielocalizer\_run-1\_bold.nii.gz & bold\_mcparams.txt
      - sub-01\_task-objectcategories\_run-1\_bold.nii.gz & bold\_mcparams ~ run 4까지
  - **cortical\_surfaces\_freesurfer**
  - **image\_space\_transformations**: 참가자(sub)/스캔(scan)별 템플릿 image, 변환은 FSL 결과. .mat 파일: FLIRT로 생성된 4x4 affine matrix. FNIRT warp files: 비선형 변환 정보.
    - ◆ code: 템플릿 이미지 생성 및 템플릿 간 변환 계산을 위한 소스 코드.
    - ◆ raw
      - phase1
      - phase2
      - structural
    - ◆ sub-01 … 21: sub-0\*/templates/ 내 각 디렉토리는 하나의 이미지 템플릿에 해당.
      - bold3Tp2: 템플릿을 포함함! group template = grpbold
        - in\_bold7Tp1: brain.nii.gz, xfm\_6dof.mat
        - in\_grpbold3Tp2: subj2tmpl\_warp.nii.gz, tmpl2subj\_warp.nii.gz 등
        - in\_t1w: brain.nii.gz, xfm\_6dof.mat
        - qa: aligned\_head\_samples.nii.gz, head\_avgstats.nii.gz 등
      - bold7Tp1: 위와 유사
      - t1w: 위와 유사 + mni152 template과 parcellation한 결과로 추정됨.
    - ◆ templates: 모든 참가자로부터 얻은 그룹템플릿 이미지(grp 접두어를 포함)
    - **retinotopic\_maps**: phase2 extension에 참여한 모든 참가자. 깜빡이는 체커보드 자극.
    - **visual\_areas**: ??

- original: Raw or minimally processed data
  - 3T\_multiresolution\_fmri: Sengupta et al.(2017)- central fixation task orientation decoding
 

이전에 7T로 수집된 데이터셋(아래아래 7T\_multiresolution\_fmri)을 3T로 보완한 데이터셋  
이전에 영화를 a-o, a-v로 모두 본 적 있는 참가자들. 7명 중 5명 중복.

    - code
    - sourcedata
    - sub-06, 09, 10, 16, 17, 20, 21
    - participants.tsv
  - 3T\_structural\_mri: Forstmann et al.(2014) - [brain structure] T1, T2, SWI, DWI
    - sub-01 ~ 21 / anat
      - sub-01\_SWI\_defacemask.nii.gz
      - sub-01\_SWI.json + SWImag.nii.gz + SWIphase.nii.gz
      - sub-01\_T1w\_defacemeask.nii.gz
      - sub-01\_T1w.json
      - sub-01\_T1w.nii.gz
      - sub-01\_T2w\_defacemask.nii.gz, json, ...
  - 7T\_multiresolution\_fmri: Hanket et al.(2014) - 기존 20명 중 6명+추가 1명 대상으로 수집  
central fixation task orientation decoding (visual cortex)  
이전에 영화를 a-o, a-v 모두 본 적 있는 참가자 7명.
    - sub-04, 06, 16, 17, 18, 20, 21
      - ses-r08, r14, r20, r30
        - func
          - sub-04\_ses-r08\_task-coverage\_bold.json, .nii.gz
          - sub-04\_ses-r08\_task-coverage\_rec-dico\_bold.json, .nii.gz
          - sub-04\_ses-r08\_task-orientation\_rec-dico\_run-01~10\_bold.nii.gz
          - ...
        - sub-04\_ses-r08\_task-orientation\_bold.json
        - sub-04\_ses-r08\_task-orientation\_rec-dico\_bold.json
    - tests
    - participants.tsv
    - 기타 recording\_motion\_physio, task-orientation, task-coverage json들...

#### phase2: 기존 20명 중 15명에 대한 추가 데이터/ movie, rtntpy, FFA lclzr, 3T eyetracking

##### 관련 링크들

- [github] studyforrest-data-phase2: An extension of studyforrest.org dataset / simultaneous fMRI/eyetracking while movie watching, plus visual localizers. a-o 이후 진행 .
  - Sengupta et al.(2016) An extension of the sf dataset for vision research, bioRxiv
    - (now published as) Sengupta et al.(2016) A sf extension, retinotopic mapping and localization of higher visual areas, Scientific Data
  - Hanke et al.(2016) Simultaneous fMRI and eye gaze recordings during prolonged natural stimulation - a sf extension, bioRxiv
    - (now published as) Hanke et al.(2016) A sf extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation, Scientific Data
  - [OpenfMRI] ds000113d
  - [OpenNeuro] ds000113: 원본, b, c, d를 통합한 데이터셋.

- code
  - rawdata\_conversion
  - stimulus
    - movie
      - compute\_segment\_cuts.py
      - cuts\_segment0~7.csv: cut\_time, 1st\_shot\_frame / 등...
    - movie\_localizer
    - retinotopic\_mapping
    - visualarea\_localizer
  - eyegaze\_utils.py
  - overlay\_gaze\_on\_video
- src
  - stimuli
    - inv\_movie\_localizer.mkv 등...: mkv 확장자는 비디오, 오디오, 자막, 이미지 등을 하나의 파일에 포함하는 멀티미디어 컨테이너 형식.
  - sub-01, 02, 03, 04, 05, 06, 09, 10, 14, 15, 16, 17, 18, 19, 20, 22~36 (N=30)
  - ses-localizer: localizer task 때 쓰는 듯. (생략)
  - ses-movie
    - func
      - sub-01\_ses-movie\_task\_movie\_run-1\_bold.json
      - sub-01\_ses-movie\_task-movie\_run-1\_bold.nii.gz
      - sub-01\_ses-movie\_task-movie\_run-1\_events.tsv: eyegaze?
      - sub-01\_ses-movie\_task-movie\_run-1\_eyelinkraw.asc.gz
      - sub-01\_ses-movie\_task-movie\_run-1\_recording-cardresp\_physio.tsv.gz
      - sub-01\_ses-movie\_task-movie\_run-1\_recording-eyegaze\_physio.tsv.gz
      - 같은 형식으로 run 1~8까지 있음. sub01은 cardresp는 없음
    - sub-01\_ses-movie\_task-movie\_bold.json
  - tests
    - bold.sh
    - eyegaze.sh
    - physio.sh
  - participants.tsv
  - 기타 Json 파일들: recording-cardresp, recording-eyegaze, (이 둘은 physio.json만) / task-movie, task-movielocalizer, task-objectcategories, task-retmapccw 등에 대하여 physio.json, bold.json이 각각 기록되어 있음.

- studyforrest-data-annotations: 영화 자막과 영화의 audio description에 대한 주석
  - code: 주석을 가져오고 변환하는 데 필요한 모든 코드
  - old: (폐기 예정) 이전에 제공되었지만, 구조가 덜 일관된 주석 데이터
  - researchcut: <Forrest Gump> 연구용 버전의 전체 클립과 타이밍이 일치하는 주석 table
    - emotions, locations, music
  - segments: <Forrest Gump> 연구용 버전의 각 segment와 타이밍이 일치하는 주석 table
    - conf(audio, visual 각각), emotions, locations, music, speech
  - src: (subdatasets referencing repositories) 사용 가능한 주석이 포함된 리포지토리
    - bodycontact, confounds, emotions, faces, locations, semanticconflict, speech
- .gitattributes
- .gitmodules
- README.md