Synchrony in Psychotherapy, Data management, example with F1044 patient data

Author: Thomas GARGOT

Abstract

Aim and Hypothesis

The aim of this project is to evaluate if it is possible to extract automatic signals of synchrony that could predict the outcomes of a psychotherapy.

Is there synchrony signals computed by SyncPy module between a patient, his parents and a therapist ? Could this synchrony signal predict outcomes of the psychotherapy ? See the full pre-registration.

INCANT Database

We began with a first database INCANT. This INCANT study aimed to evaluate the efficacy of the Multi Dimensional Family Therapy (MDFT) for cannabis use disorders in adolescents. Some patients received treatment as usual and others MDFT. This database come from a large european psychotherapy study. The main outcome was the cannabis consumption evaluated by the Timeline Follow-Back (TFLB) questionnaire.

For a pilot study, we focused on the analysis of the F1044 subject, his family (father and mother) and the therapist. We saw that the manual extraction of speech was very laborious and quite subjective so difficult to replicate. We saw that there was a lot of modules that enable automatic social signals. We decided to extract automatic signal and began to focus on motion history.

Unfortunately, we couldn't answer to this question since the signal extracted from this videos was not very good. It was difficult neither to categorize easily the patients in good or bad responders neither to detect specific periods of the psychotherapy that could be of particular very good or very bad quality. Furthermore, it was difficult to compare synchrony at different moments since the configuration of the participants could be very different.

Monrado Database

Consequently, we switched to the Monrado database, an other video database, with a better video quality and experimental design. It evaluates the difference of behavior in a situation of conflict vs no-conflict.

The main psychometric data in this database were attachment scores, i.e. a evaluation of the relationship of the subject with other people.

Poster

Extracting automatically social signals from psychotherapy sessions

Thomas Gargot[1,2,*], Giovana Varni[2], Jonathan Aigrain[2],
Mohamed Chetouani[2], Michel Spodenkiewicz[1,2], David Cohen[1,2]
1. Hôpital de la Pitié Salpêtrière, 2. Institut des Systèmes Intelligents et de la Robotique, Paris, France
* thomas_gargot@hotmail.com

Introduction and motivation

Psychotherapy is an important part of treatment of mental disorders alone or complementary with pharmacotherapy. Some techniques are now widely evidence-based and very cost effective (Layard & Clark, 2014).

Most of the studies are indirectly based on patient reported outcomes or problematic behaviors that are evaluated before and after the psychotherapy. Unfortunately, studies hardly control what is directly happening during psychotherapy, especially the interaction between the patient and the therapist that could be a predictor of the psychotherapy efficacy. Consequently, it is difficult to make precise links between theory and practice, control its application and understand which of its ingredients are the most important.

It is yet possible to annotate manually videos. However, this task is challenging since it can be either very repetitive (annotation of turn-taking or non-verbal behavior) or very technical (annotation of application of some specific techniques like in motivational interviewing (Moyers et al, 2015).

In the future, an automatic feedback of the psychotherapy could also even help a therapist, to guide him to reorganize a treatment.

Towards an automated research framework

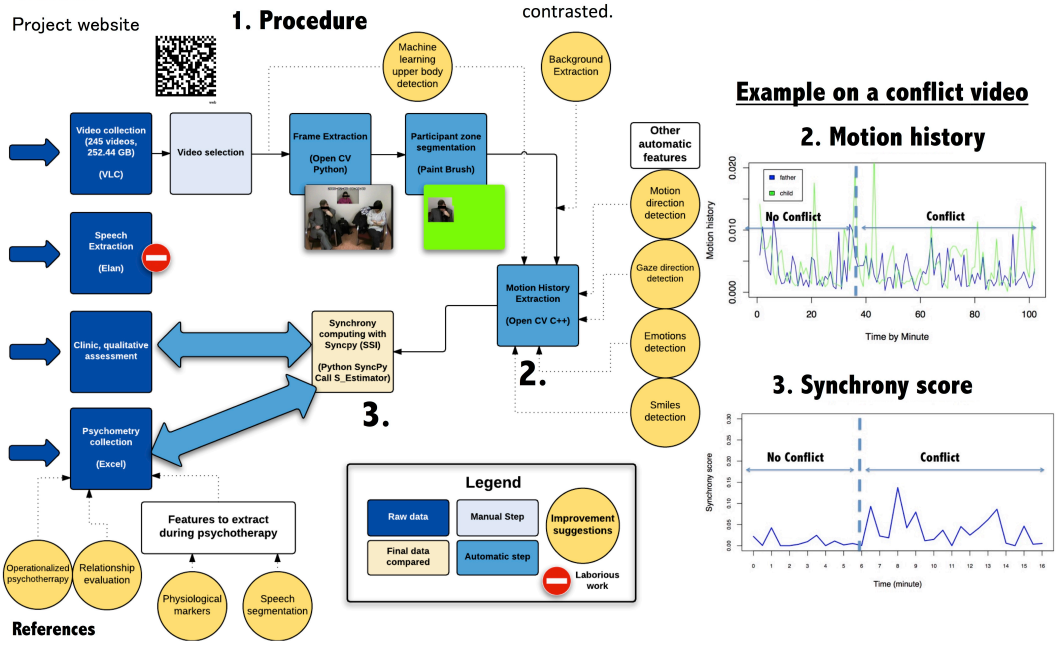
Here we suggest a research framework (Fig 1.) to extract automatically social signals from psychotherapy videos. We focused on motor synchrony since it was found to be a predictor of psychotherapy outcome in a first study (Ramseyer, 2011) and a relevant signal on the study of mother-child interaction. Varni et al. 2015 developped the SyncPy open source python library to help researchers and practitioners to automatically analyse synchrony. It could be possible to measure synchrony even in familial therapies. Other features could be quite easily extracted manually like direction of the body, gaze direction or smiles. Here, we show an example of a video before and during a conflict between a girl and her father (Fig 2. and 3.).

Using an interdiciplinary and open science approach, we developed some modules toward this goal that could be freely and easily re-used by other teams in other databases or with other modules.

Future work

The next step would be to design a specific database anticipating some technical problems (speech segmentation, overlapping of subjects, micro-movements) and define more precisely different conditions of psychotherapy that could be contrasted.

Project website



References

Layard, R., & Clark, D. M. (2014). *Thrive: The power of evidence-based psychological therapies*. Penguin UK.

Moyers, T. B., J. K. Manuel, and D. Ernst. "Motivational Interviewing Treatment Integrity Coding Manual 4.1." (2014).

Ramseyer, Fabian, et Wolfgang Tschacher. « Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome ». *Journal of Consulting and Clinical Psychology* 79.3 (2011): 284-295.

Varni, G., et al. "SyncPy: a Unified Open-source Analytic Library for Synchrony." *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And influence*. ACM, 2015.

No potential conflicts of interests

Originality declaration

This study is the second study to evaluate the relevance of synchrony as marker of the

quality of psychotherapy after the study of Ramseyer et Tschacher, 2011. It is the first to study it in families, which both configuration and theorical frameworks are different. The package used for synchrony was developed by the engineers of the lab and it was the first time that it was used in real world data. The input of psychiatrist and psychotherapist helped to understand the kind of signals that could be extracted and relevant from psychotherapy videos.

It is the first study to use an open science approach among the 3 studies (Ramseyer et Tschacher, 2011; Orsucci et al., 2016) that used synchrony as predictor of quality of relations since raw data extracted from videos, analysis reports and preregistration are published on the website project.

Collaborations

In alphabetical order :

Name	Function
Jonathan Aigrain	Engineer, Creating frames from videos , extracting Momentum from videos frame by frame, knowledges in Python, especially OpenCV
Nicolas Bodeau	Bio-statistician and informatician in ISIR and La Pitié, supervising for R
Mohammed Chetouani	Engineer, supervising the project, and the SyncPy library
David Cohen	Psychiatrist, Original idea, supervising the project
Catherine Saint Georges	Psychiatrist, knowledge in interpreting synchrony scores in psychiatry
David Reversat	Engineer, Debugging of SyncPy
Michel Spodenkievitch	Psychiatrist, Collection of the database (videos from INCANT and Monrado studies), direct supervision
Giovanna Varni	Engineer, Development of the SyncPy library

The Monrado and Incant groups recruited subjects and psychometric data

I helped to define and operationalize the scientific question from the general idea, the

tools and databases available. I did bibliography based on the team work and Ramseyer article. I cleaned and prepared the databases. I adapt the Python and R script to make the extraction of the signal and perform the analysis. I realized the figures and diagrams to explain the process of the method and describe the results.

Why and How evaluate psychotherapies?

Psychotherapy evaluation is quite difficult since it must evaluate the basics of human to human relationship. A lot of theories developed in this field (psychoanalytical, systemic, behaviorist, hypnosis) with very different practices and level of evaluation. There is not for the moment very clear and homogenous framework of its efficacy. Relationship between the

Potential role of cognitive sciences in improvement of psychotherapies

- Cognitive therapies
- Learning theory
- Attention
- Cognitive dissonance and engagement theory
- Bandura
- Miror neurons
- Therapeutic alliance (Dodo Effect / Thrive book)

Theorical advantages of synchrony studies

Data structure

The data consists of 2 databases with for each some videos and psychometric data.

INCANT study

Study protocol

The Incant study aimed to evaluate the efficacy of Multi Dimensional Family Therapy (MDFT) a familial therapy with other treatments Treatments As Usual (TAU). Videos were recorded to check if the therapist was applying the psychotherapy that was assigned to each family.

The MDFT is based on several general principles likes topics that needs to be addressed,

attitude of the therapist.

Videos

- We collected a 252.44Gb database of 277 Videos, with a rate of 25 frames by second.
- They are encoded in VOB format.

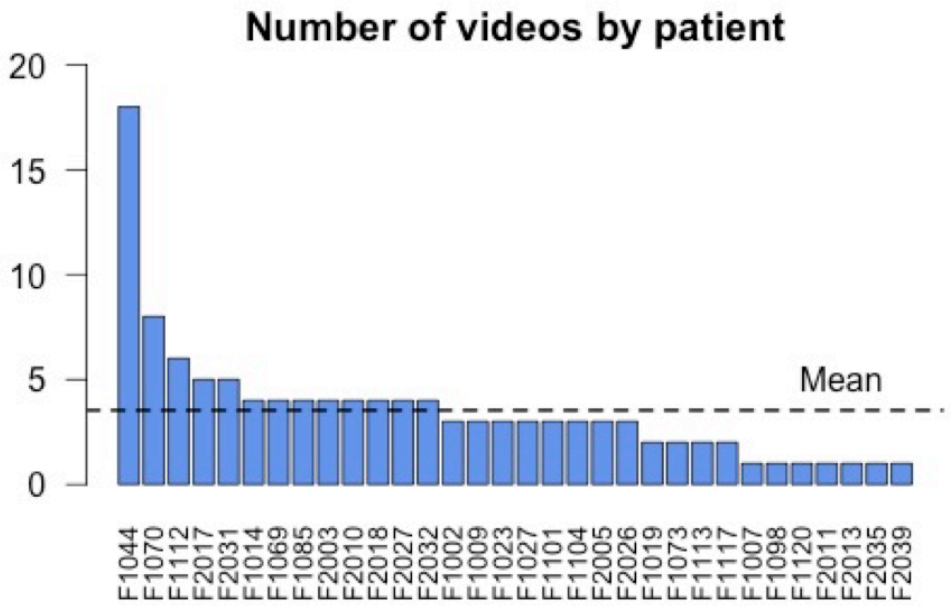
Psychometric data

- It consist of 3 excel files with :
  - the Timeline Follow-Back (TLFB)
  - the Youth Self-Report syndrome structure (YSR)
  - the Child Behavior Checklist syndrome constructs (CBCL).

However, we didn't have any data dictionary with the definition of all variables and the exact questionnaires used to collect this information, the coding of Non Available data that may be -99. We planned to meet the team that wasn't unfortunately available at that moment. A lot of videos were lost or not recorded. We couldn't know from this database which patient received MDFT or TAU. It was organized in a European level. We collected only the video and psychometric data from the 2 French centers.

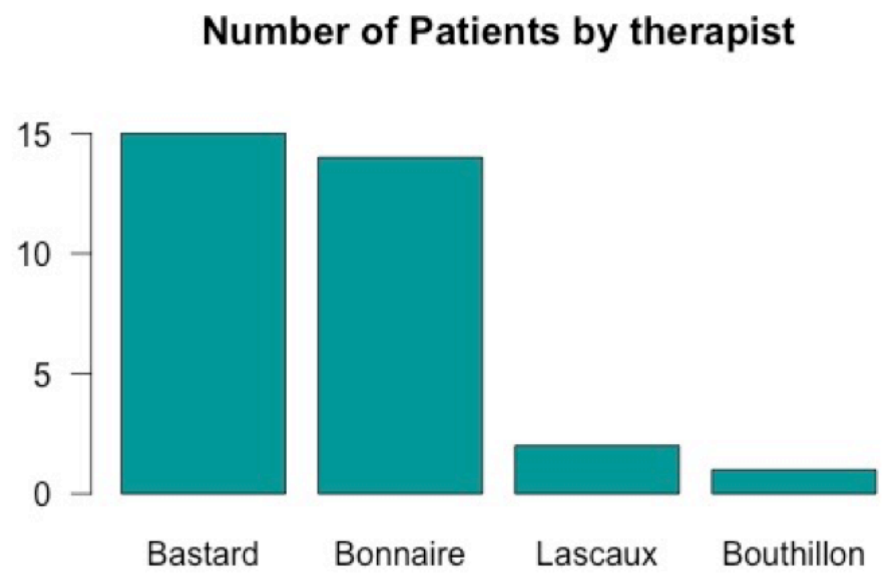
Pilot study

We decided to make a pilot study on the most complete subject : F1044 since we have 18 videos with him or his family.



We can ask ourselves if this subject representative:

- The therapist is the first most representative concerning the number of patients and the first concerning the number of patient videos.



- The patient is a male like in 93 % of videos. The therapist is a female like in 99 % of videos.
- Two centers were included. This patient come from the main center.
- We have all our psychometrics evaluations (cannabis consumption : initial, 3m, 6m, 9m, 12m).

Unfortunatelly, the first, second, 10th and 11th videos were lost.

Nomenclature

F1044 is the name of the subject studied (called *patient*). When referring to several persons, variables take the name of the shorter form *pa* for *patient*.

He has a *mother* (*mo*) a *father* (*fa*) helped by a *therapist* (*th*). When a variable is referring to several participants, it is organised in alphabetical order separated by underscores, eg. *SSI_fa_mo* refers to the synchrony index (*SSI*) between the *father* and the *mother*. *SSI_mo_fa* doesn't exist.

This familly had several consultations with the psychotherapist (*therapist*). Some of them were video recorded. These videos are names with the name of the subject + an index letter. They can subdivised after that with numbers (eg *F1044C*).

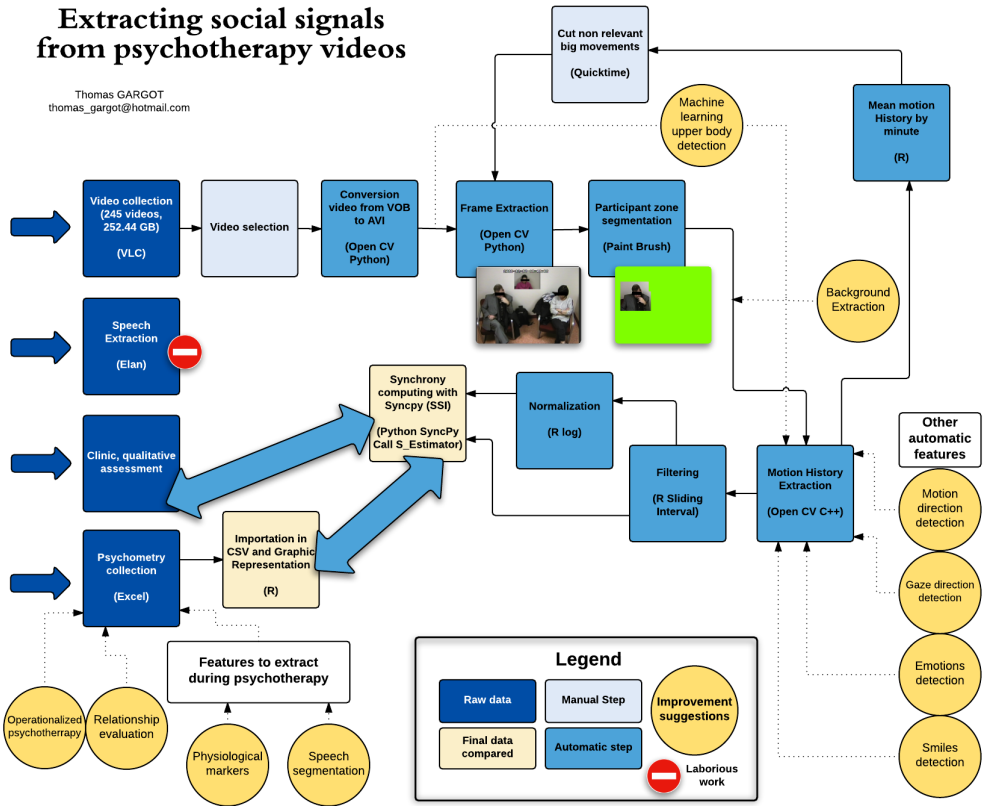
Softwares used

- Excel tables manager
- R Statistical software
    - xlsx package
- Python Programing language
    - opencv Module for extracting information from videos
    - pandas Module for managing data frame
    - matplotlib Module for making plots
    - syncpy Module developed by the team for studying synchrony
    - FFMPEG Module for converting videos
- C++ Programing language used to extract motion history
- Paintbrush Drawing program used for the frame segmentation
- Quicktime Video player used to cut videos
- VLC Video player
- Elan Annotation software use for the essay of annotation
- Mou : Markdown editor
- Git
    - GitHub website public version control repository

Steps

Description of the database

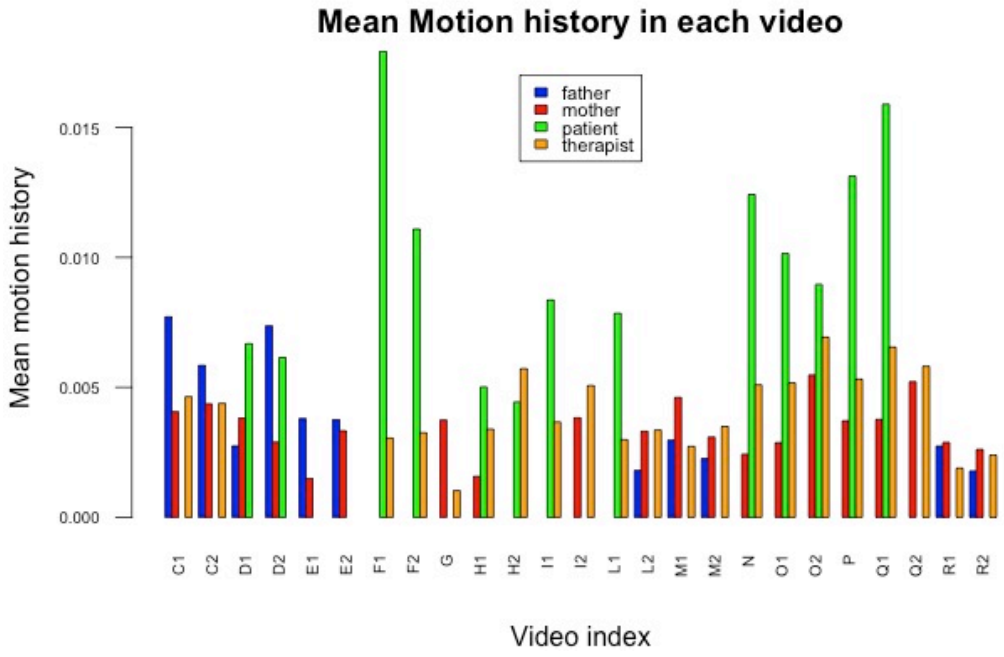
Summary figure of technical steps



Video Configurations

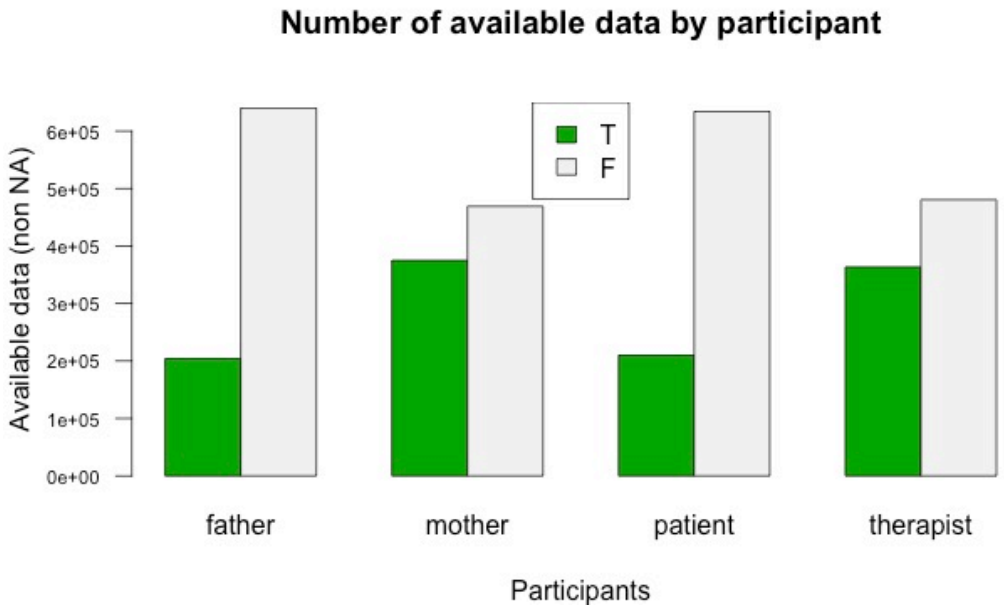
Spatial organisation

In the first center, psychotherapy sessions consisted of filmed familial psychotherapies with 2 to 5 peoples organized in a circle. Two participants of the psychotherapy were filmed from the front. Another subject (often the therapist) is filmed from the other side. Its video is embedded in a window. We can notice the date displayed on the video. Sometimes, there is an overlap of different subject in the same place.



In this barplot of the mean motion history for each subject for each video, we can see that the configurations of the participants involved for the F1044 subject can be very different between videos. There is between 2 and 3 participants. There is never the 4 participants involved. In some videos, we don't have any information from the therapist because she isn't filmed. The patient is not very present in the therapy (because of compliance).

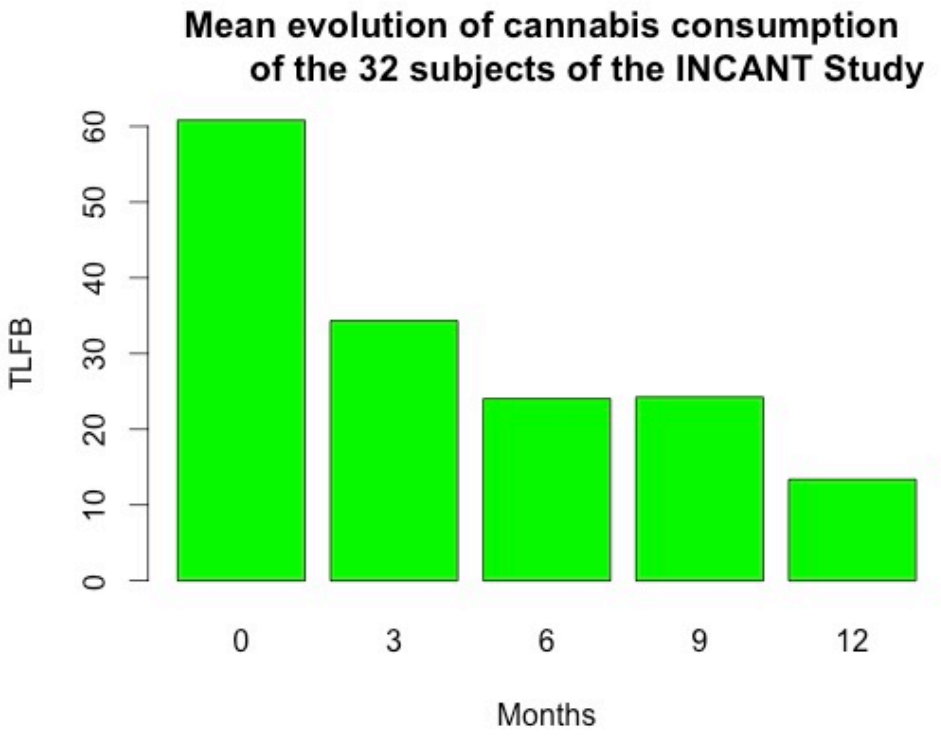
When we sum up the data we get, we can see that the mother and the therapist are the most present participants.



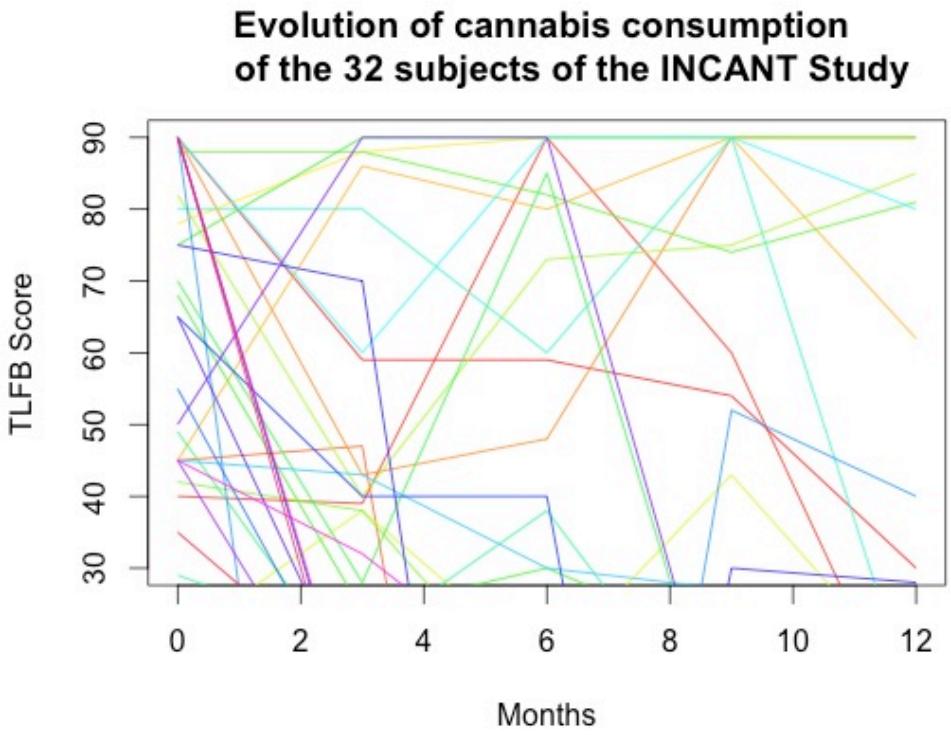
Psychometric data

Files were collected in xls format. They were imported via the XLSToCSVConvertor.R R script. It returns a dataINCANT.csv csv file and a dataCannabis.csv file.

Globally, the cannabis consumption is decreasing with time for the subjects.



We didn't have any group defined. When we look of the evolution of the cannabis consumption, it is not very easy to define good and bad responders.

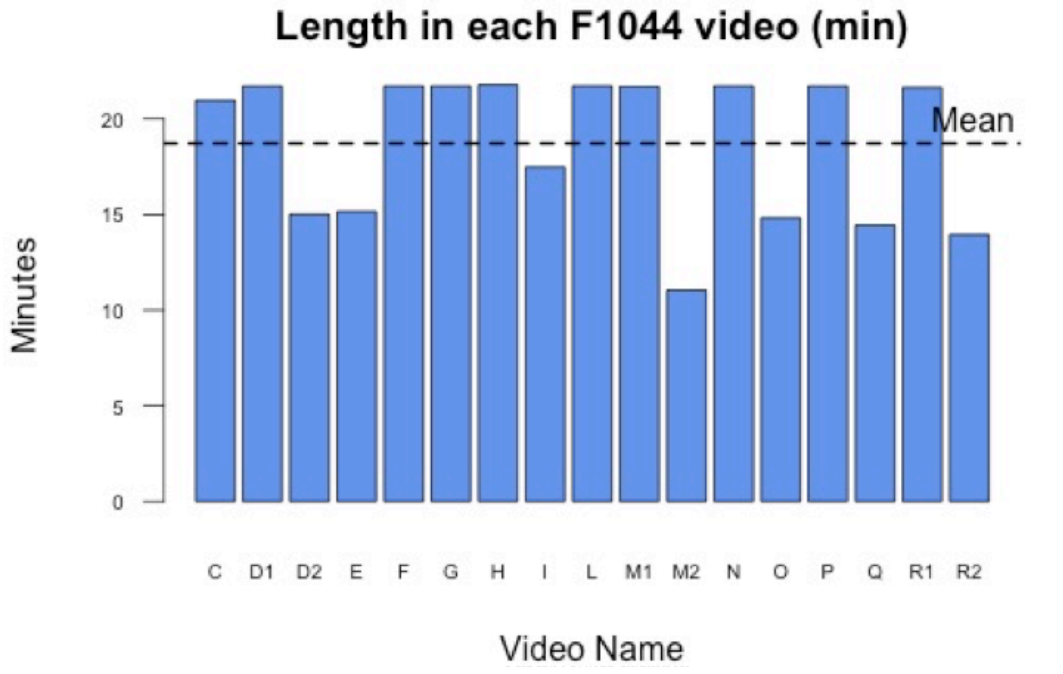


All the evolutions of this scores are plotted here.

Time organization

Video length could were very different because it was necessary to reorganize them.

However, we couldn't use the raw database. It was necessary to cut the videos with Quicktime software since a phone was used to communicate between the main psychotherapist and the supervisor and the therapy stopped at that moment. Sometimes, the participants made some pauses or the same video was used for different patient.



Furthermore configurations could change (eg patient-mother-therapist then mother-therapist) during the psychotherapy,

Raw data

The first step was to extract frames with the Frames_extractor.py python script made by Jonathan Aigrain with FFMPEG.

Idea	
	In a lot of other domains of fundamental algorithm development, computer labs compete on public databases to improve the detection rates of some features for instance. This work are presented during challenge sessions in congresses. However, there is a problem with confidentiality when we deal with psychotherapy videos.
	It would be better to make a specific database for this kind of automatic analyses to avoid changes of configuration of the room, the luminosity, the place of the different subjects, the number of subjects, the movement of the camera during recording.

A. Motion extraction

1. Extracting frame from the beginning of the video

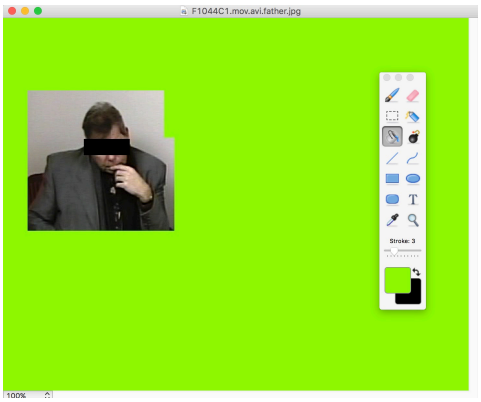
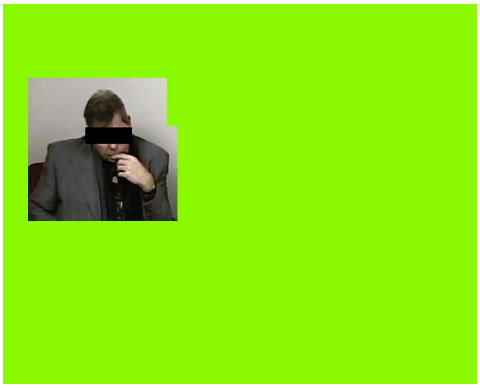
Extract a frame with *Frames_extractor.py* with a python script from the beginning of the video to make a mask. This script was developed by Jonathan Aigrain with Open CV.

Here is an example of the father (eyes blurred to keep anonymity).

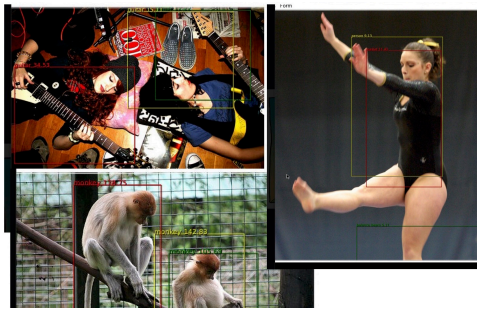
Idea	
	It was suggested to take a mean image of the video to be more precise instead of taking a video from the beginning.
Beware	
	Notice the date that was always masked after during the step 2.
	We can change the value of the median filter.

2. Make a mask for each relevant part of each video with Paintbrush

Select a part of the video with each participant (father, mother, patient, therapist) with *Paintbrush*. Take only the upper part of the body to compare them since you can see only the upper part of the therapist.

Paint masking	Result
	

This operation is repeated for each participant on each video.

Beware	
	It is necessary to always use the same background color, for instance RGB green. (0,255,0)
	Participants can move during the video (change their seats, leave the room). It would be preferable to anticipate it. Camera is moving sometimes too. Mean Motion history by minute can help us to detect big changes or disappearance of a participant.
	Participants are labelled in the mask name (eg *F1044C1.avi.father.jpg*)
Idea	
	Deep learning software or machine learning used for categorization/detection of people could make this process completely automatic. 

3. Motion history extracted

The VOB videos were converted in AVI format in the form *F1044C1.avi*.

It is extracted with the video and the mask with a C++ script. It returns CSV file of the form "*F1044C.VOB_res2.csv*". See files here.

Headers :

- **frame** : number of the frame (original rate of 25 frames by second)
- **father** : motion history of the father frame by frame. It ranges from 0 (same frame than the previous frame, no movement) to 1 (every pixel change).
- **mother**, : idem for mother
- **patient** : idem for patient
- **therapist** : idem for therapist
- **file** : name of the file in the form *F1044C.VOB*

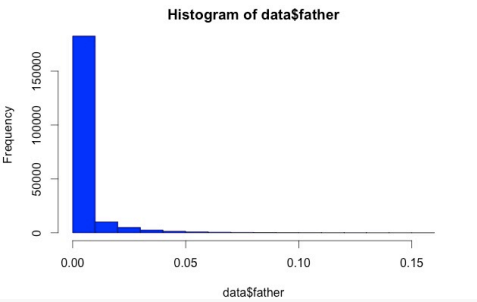
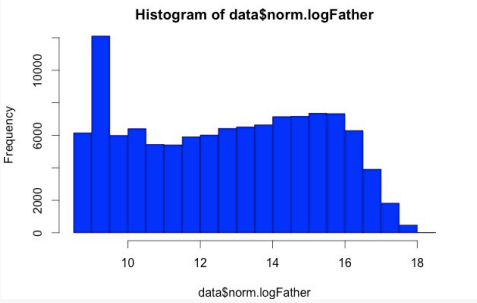
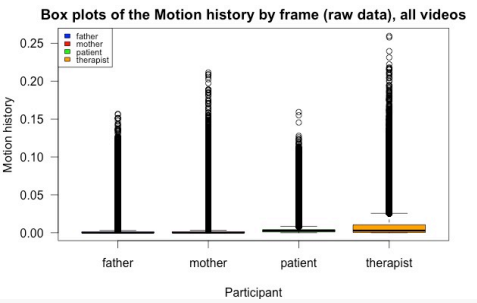
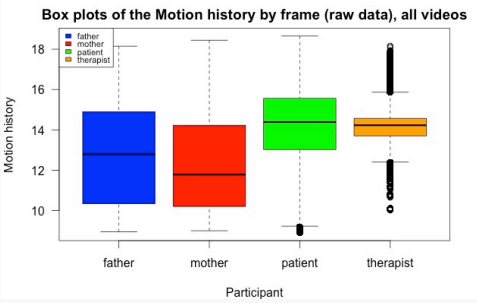
- **NA** : (Non available) corresponding to absent subjects (NA column) or bad quality of frames at the beginning or end of the videos making impossible extraction of motion history for any participant (NA line). After a problem of encoding, from .VOB, to .mov (video format exported from quicktime when cut are made) and then to AVI, there is a problem of encoding and the length of the video is not well recognized and NA lines are generated that correspond to nothing.

If we plot (in histograms and box plots) this taw data (1st column), we notice that the distribution is not normal at all. Very small motion history values are over represented and and bigger motion history are much more rare with a very long tail.

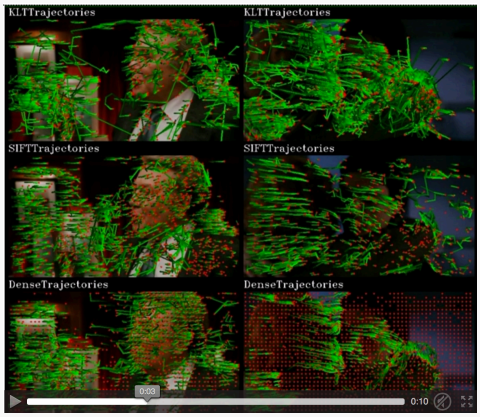
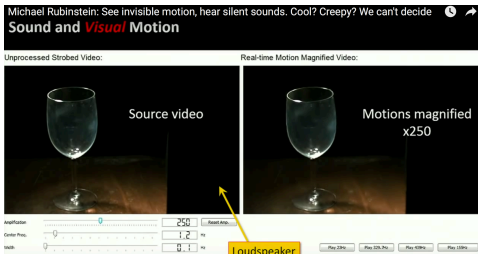
To normalize the distribution to compute synchrony scores on it, we made the natural logarithm. It produces negative numbers. SyncPy can't compute negatives scores, they are so shifted to positives values with an arbitrary value of 20 to avoid to keep extreme negative values.

Values equal to 0 can't be loged. They generate a -Inf value. These values are set to NA. We lose the information of no movement at all. If we give a arbitrary value to this data (eg, the minimum value, they are over represented).

This normalized log data is shown on column 2.

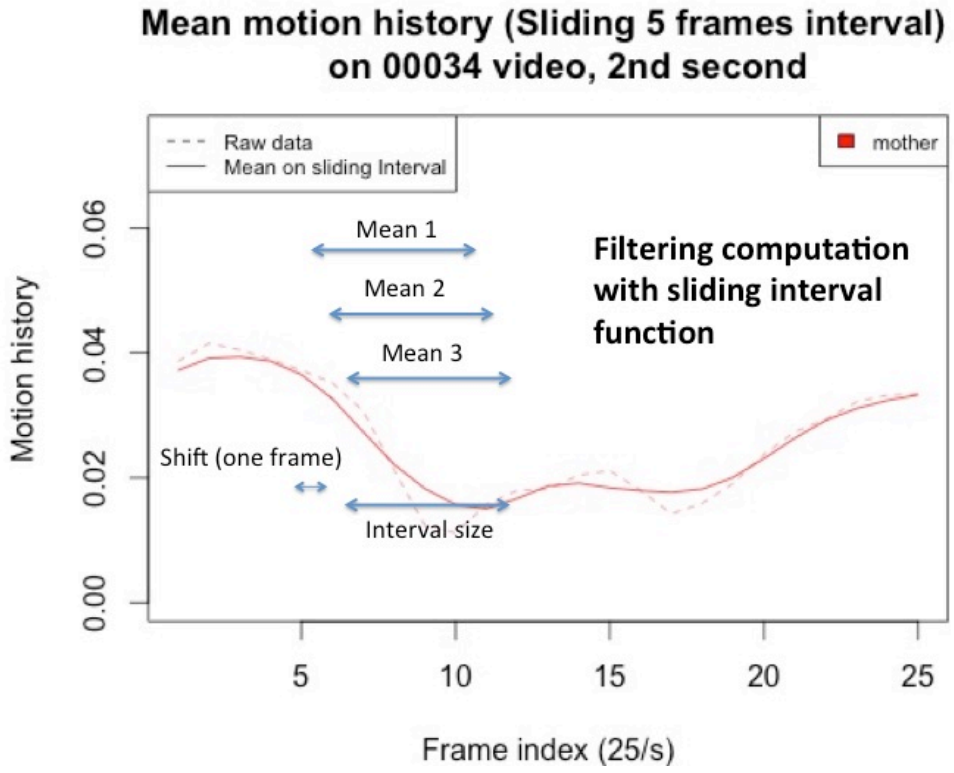
Raw Motion history	Natural Log Motion history
Father	
	
All participants	
	

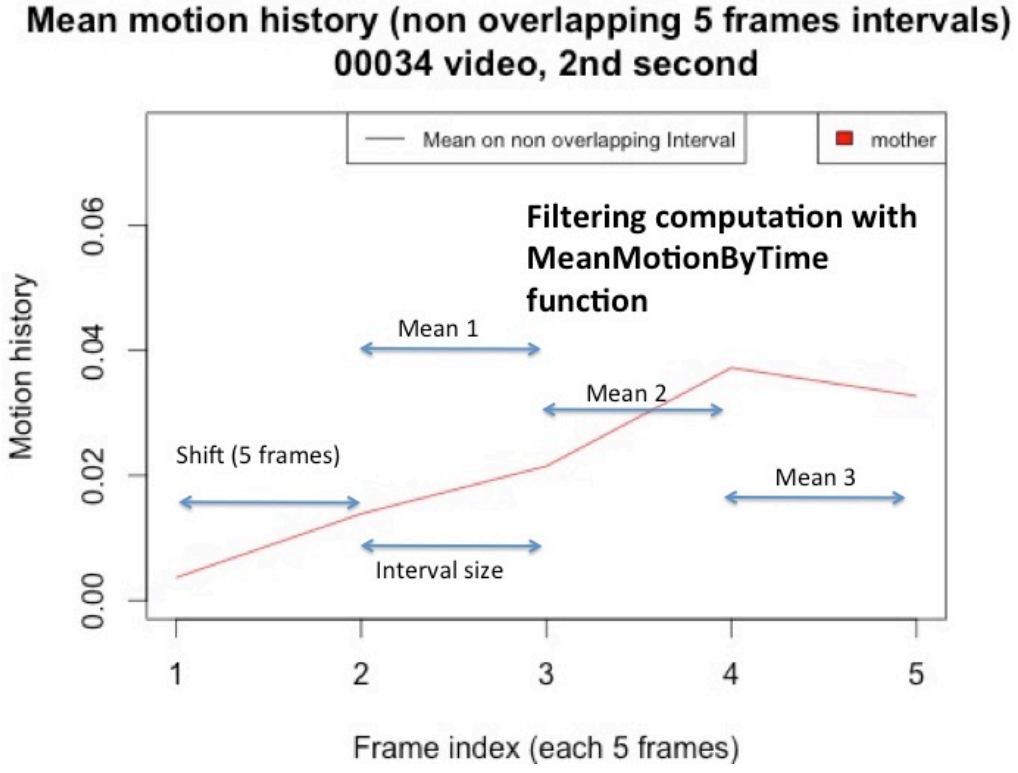
If we check all participants, we can see that the father and the mother motion history distribution is very similar. However, the therapist, which is always in a small window of the video, as a very different distribution. We have less signal on it. In some videos the patient is in this window, it explains, it intermediates position.

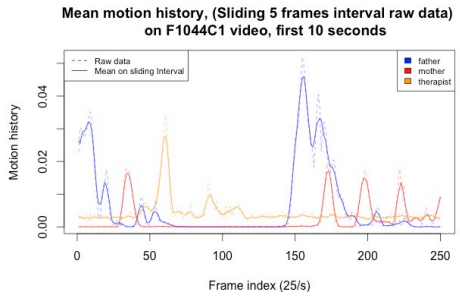
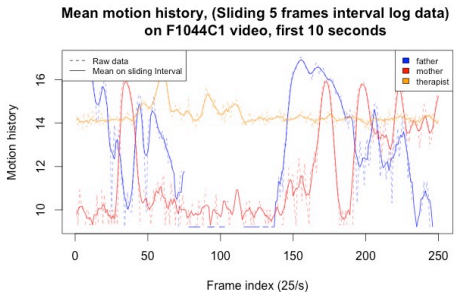
Idea	
	- It could be possible to extract background to improve the quality of the motion history. See for instance backgroundExtractor.py with openCV BGS library.
	- It could be interesting to evaluate the trajectories of the people. If they are going toward the same direction or not. For that, the optic flow modules of open CV or the Dense Trajectories Video Description could help. 
	Tiny changes can't be always obvious. Some softwares can amplify a lot very tiny changes in specific frequencies that can't be seen normally with naked eyes. 

4. Motion history filtered with slindingInterval function

Motion history raw data is filters with a R function on Rmarkdown script analysis. It generates CSV file of the form "F1044C.VOB.slideddata.csv". See files here.


Mean motion history (Sliding 5 frames interval)
on 00034 video, 2nd second
Raw data
Mean on sliding Interval
mother
Mean 1
Mean 2
Mean 3
Filtering computation
with sliding interval
function
Shift (one frame)
Interval size
Motion history
0.00
0.02
0.04
0.06
5
10
15
20
25
Frame index (25/s)


Mean motion history (non overlapping 5 frames intervals)
00034 video, 2nd second
Mean on non overlapping Interval
mother
Filtering computation with
MeanMotionByTime
function
Mean 1
Mean 2
Mean 3
Shift (5 frames)
Interval size
Motion history
0.00
0.02
0.04
0.06
1
2
3
4
5
Frame index (each 5 frames)

Sliding interval (raw Data)	Sliding interval (log Data)
	

Headers :

- **video** : name of the video (eg F1044C.VOB)
- **frame_index** : number of the frame (original rate of 25 frames by second)
- **slidedFather** : filtered motion history with the sliding motion history function making mean on 5 frames (2 frames before and after the index frame). It ranges from 0 (same frame than the previous frame, no movement) to 1 (every pixel change).
- **slidedMother** : idem for mother
- **slidedTherapist** : idem for therapist
- **slidedPatient** : idem for patient
- **NA** : corresponding to absent subject

Idea	
	It is possible to change the size of interval to change the ration signal/noise
Beware	
	Bad quality of first frames at the beginning or end of the videos with NA were deleted to prepare the next script analysis.
	The number of frames is changed since it is necessary to get frames before and after the index.

4. Computing synchrony score with this filtered motion data

Amplitude (difference between min and max) and baseline levels (min) are not relevant to compute synchrony.

The script in Python, Call_S_Estimator.py takes this filtered motion data and compute a synchrony score for each association possible of two or more subjects.

Download and install SyncPy.

The script Call_S_Estimator.py must be installed in Syncpy installation in examples folder. Create a folder *SynchronyCSV* in this folder.

You need to specify,

- the folder where you put your data (eg dataFolder = '/Users/Ofix/Documents/Fac/internat/Recherche/projets/synchro/synchroData/CSV/filtered/')
- the list of videos you want to analyze, (eg VideoList = ["F1044C.VOB", "F1044D1.VOB", "F1044D2.VOB", "F1044E.VOB", "F1044F.VOB", "F1044G.VOB", "F1044H.VOB", "F1044I.VOB", "F1044L.VOB", "F1044M1.VOB", "F1044M2.VOB", "F1044N.VOB", "F1044O.VOB", "F1044P.VOB", "F1044Q.VOB", "F1044R1.VOB", "F1044R2.VOB"])
- the number of frames used in an interval to compute synchrony, Nota bene, there are 25 frames by second. (eg. numberOfFramesByInterval = 25*10)

This process is relatively long and can take several minutes, especially with short intervals.

It returns synchrony CSV files that can be found here.

Headers :

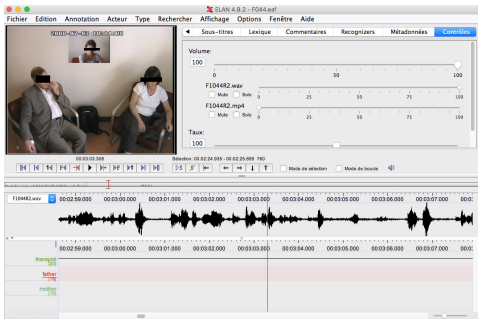
- **Interval**
- **Time_min**
- **video** : name of the video (eg F1044C.VOB)
- **SSI_fa_mo** : synchrony index between the filtered motion history of the *father* and *mother*. It ranges from 0 (no synchrony at al, not any relationship between the different participants of the score) to 1 (complete synchrony, the different elements move a lot at the same moment or don't move at all at the same moment).
- **SSI_fa_mo_th** : idem between *father*, *mother* and *therapist*
- **SSI_fa_th** : idem between *father* and *therapist*
- **SSI_mo_th** : idem between *mother* and *therapist*

There is not any NA in this file. Only the possible combinations are computed (eg there is not the *patient* (*pa*) in the video so, there isn't *SSI_fa_pa* signal neither any other combination with the patient).

B. Speech extraction

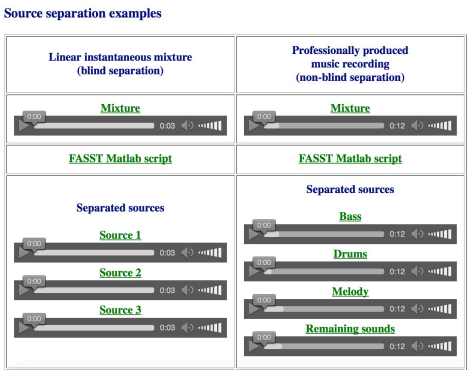
1. Manually

It is possible to use the software Elan to annotate a video. You need to extract first the sound (with Free Video Converter for instance) to get a waveform and import the video and the waveform.



Then you can annotate when someone is speaking. This process is very fastidious (1 week work for one hour video at the beginning and so was gave up).

2. Automatically

Idea	
	Some softwares could be used to make this process easier. The FASST software get very good results to segment a music sample in very different instruments. However, different voices are too close to distinguish them easily automatically. 

We can see that the evolution of cannabis use by F1044 is not straightforward. It is quite high at the beginning, decreases and then increases. Consequently, we decided to select 4 subjects with a clear decrease of cannabis consumption.

- F1002, patient, 3 videos
- F1073, patient, 2 videos
- F1069, patiente, 4 videos
- F1101, patient, 3 videos

All of them with the same therapist Mrs Bastard

Advantages and limits of this database

Advantages

INCANT study was done internationally with:

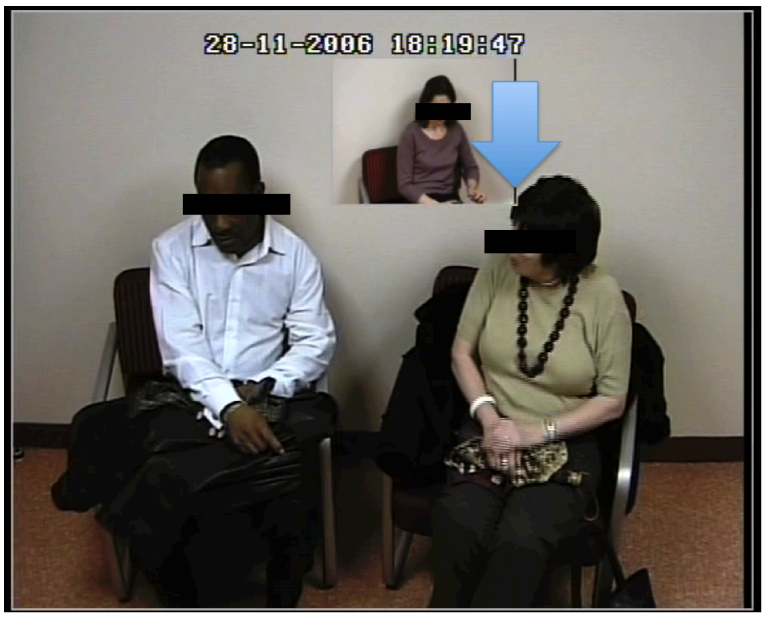
- good level of scientific method with a pre-registration done in major journals in a psychotherapy field where the level of quality of the studies can be not very good.
- potential possibility to retest the model on videos taken from other countries if available and/or generalize it.
- Data came from almost real life family therapy settings the would be easier to generalize to everyday practice.
- Pragmatically, it was the first database we got, so we could begin to ask real life problems.

Drawbacks

Experimental set-up

It is however not very defined in the articles. Consequently, it was difficult to contrast synchrony between two well defined periods. We could suggest to have a much more defined psychotherapy and even before studying a psychotherapy to study specific situations : open vs closed questions,

Technical problems



In this frame, we can notice that there is an overlap between the region of interest of the therapist (up, embedded in a window) and the mother (right). Some motion history from the mother could consequently be lost.

In this frame, we can see that half of the motion history from the mother couldn't be recorded since it was outside the camera view. Moreover, there is an overlap of the date with the patient frame (top). We can notice too that this window frame .

In this frame, we can see that the date is overlapping with the therapist.

Monrado study

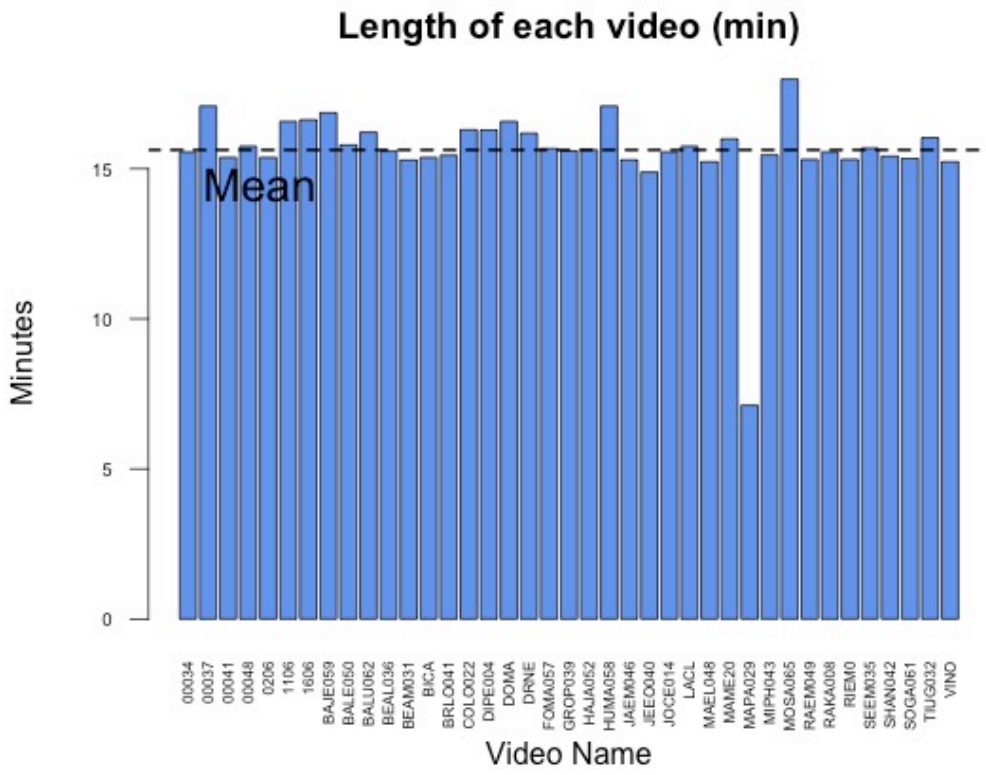
- We collected a database of 85.48 Gb of 40 videos, with a rate of 25 frames by second.
- They are encoded in a MTS format in a much better quality than the previous database.

Psychometric data

- It consist of one excel file with:
  - Identification number
  - Demographic information
  - Attachment style
  - TAS Score
  - STAIYA Score
  - BDI score

Description of the database

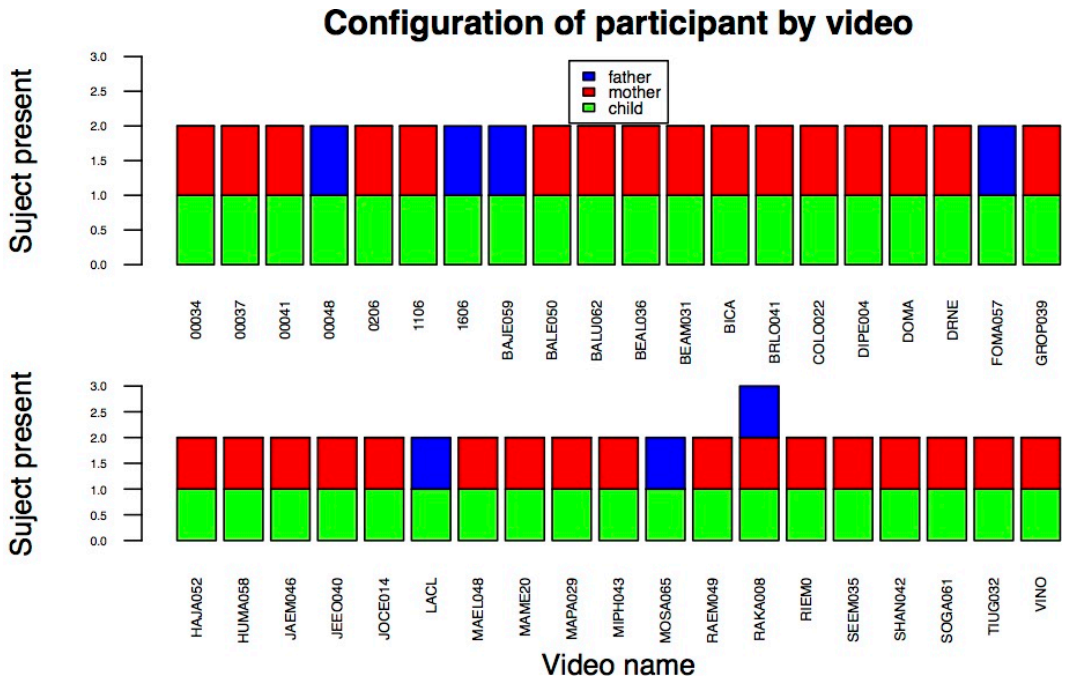
Length of the videos



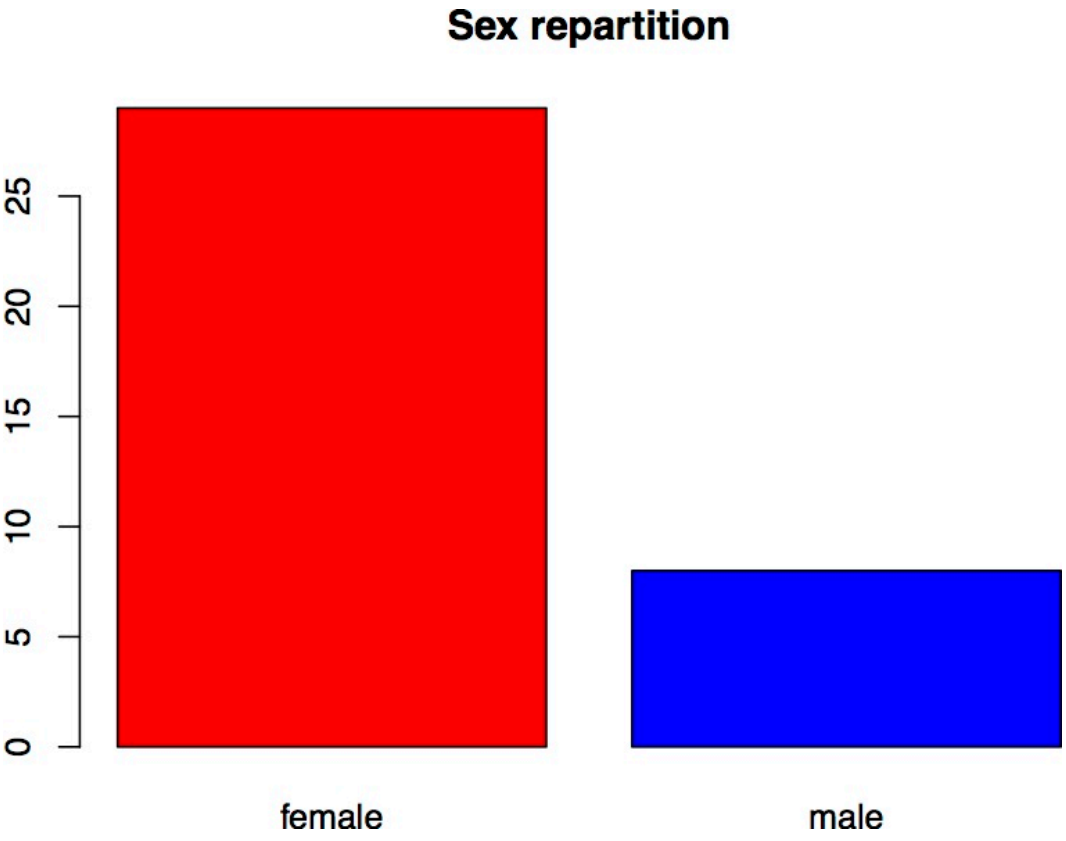
We can see that the videos are very similar.

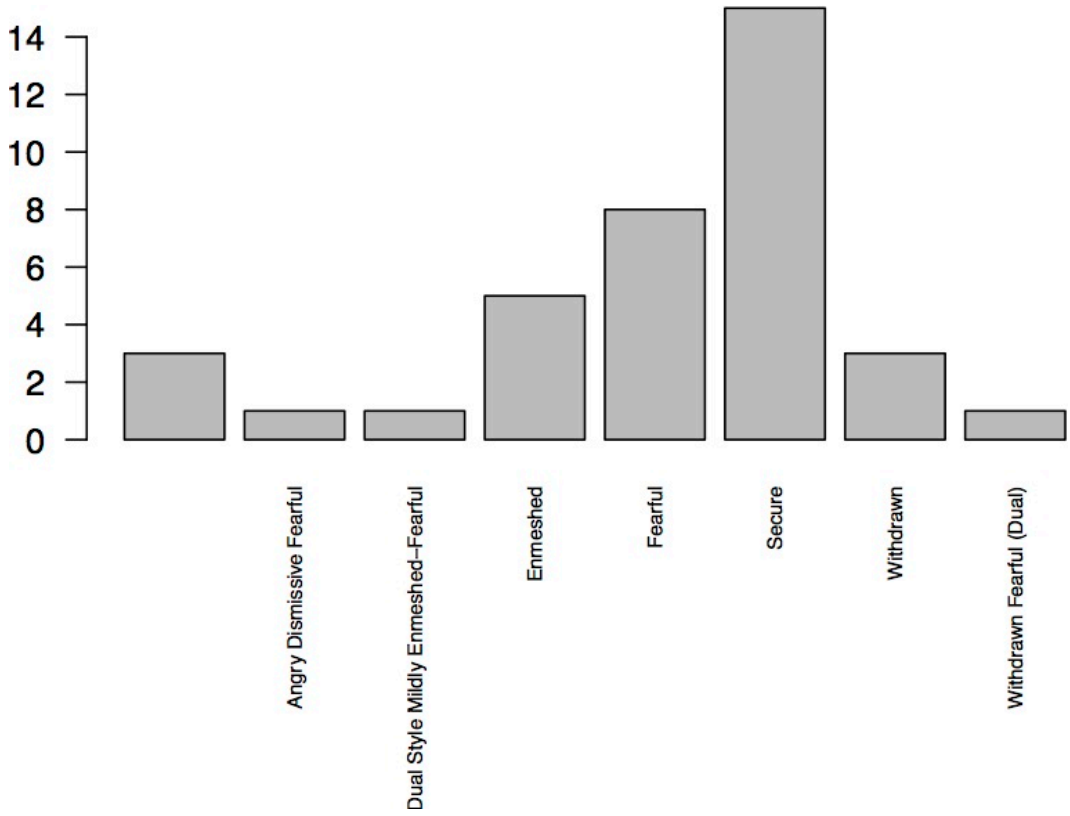
Configuration

The participants are two in front of the camera. We can see that the quality of the video is much better. There is not overlap between the subject. There isn't embedded subjects. There is not any date embedded.



The configuration in the different videos are very similar. There is only one video with 3 persons.





Technical problems

None

Results

Discussions

It could be relevant to annotate the level of conflict by blinded rotator since the conflict is not always completely obvious nor strong. We could extract other features :

- Hormons
    - cortisol
    - ocytocin
- Cardiac Frequency : direct or indirect on videos with systems like Lambda or webcam pulse detector
- Anotation of conflict and non conflict period with Elan software
- Electrodermal activity that seems to be relevant in stress, dissonance cognitive paradigms. Its history as feedback of psychotherapy is ver old since it was used in psychoanalysis by C. G. Jung

- Motion capture Optitrack
- Kinect
- Source separation source : FANNS