

# How synchrony detected from videos could predict the quality of a therapeutic or parent relationship

Author: Thomas GARGOT

Synchrony articles before

## How synchrony detected from videos could predict the quality of a therapeutic or parent relationship

Author: Thomas GARGOT

Abstract

Introduction and motivations

International Cannabis Need of Treatment (INCANT) study Database

Monrado Database

Conclusion

Poster

Originality declaration

Collaborations

Theoretical background

Why and How evaluate psychotherapies?

Potential role of cognitive sciences in improvement of psychotherapies

Learning theory

Cognitive neuroscience and education

Cognitive dissonance and engagement theory

Bandura

Mirror neurons, mirroring and empathy

Theoretical advantages and pitfalls of synchrony studies

Open science, Open data, Confidentiality

INCANT Study

Introduction

Cannabis addiction

Cannabis addiction treatments

MDFT and the INCANT study

INCANT Study in France

Material and methods

Videos

Softwares used

Psychometric data

Results

Pilot study: F1044 subject, his father, mother and therapist

Description of the database

Different steps of the social signal extraction

Verbal signal extraction to compute verbal synchrony

Spatial organisation

Time organization

Raw data

A. Motion extraction

1. Extracting frame from the beginning of the video

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

Technical problems

3. Motion history extracted

Headers :

4. Motion history filtered with slidingInterval function

Headers :

4. Computing synchrony score with this filtered motion data

Headers :

Advantages of this database

Limits of this database

Experimental set-up

Technical limits

Conclusion

MONRADO study

Aim of the study

Configuration

Selection of relevant videos

Length of the videos

Psychometric data

Attachment style

TAS Score

STAIYA Score

Description of the database

Extraction of motion history

Technical problems

Results

Discussions

Discussion

Experimental discussion

**Perspectives**

**Bibliographie**

## Abstract

Introduction and motivations

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 actually happening during psychotherapy, especially the interaction between the patient and the therapist that is 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.

Here we suggest a research framework to extract automatically social signals from psychotherapy videos and tested it on two databases. We focused on the extraction of synchrony of the motor signal since it was considered as a predictor of psychotherapy outcome in a first study (Ramseyer, 2011) and a relevant signal for the study of mother-child interactions. We adapted a python software in open source [SyncPy](#) (Varni, Avril, Usta, & Chetouani, 2015) to compute this synchrony.

### International Cannabis Need of Treatment (INCANT) study Database

We began with the International Cannabis Need of Treatment (INCANT) study database from France. This [INCANT](#) study aimed to evaluate the efficacy of the [Multi Dimensional Family Therapy](#) (MDFT) for cannabis use disorders in adolescents.

For a pilot study, we focused on the analysis of the F1044 family and the therapist. Unfortunately, we couldn't answer to this question since the signal extracted from this videos was not good enough. Furthermore, it was difficult to categorize easily the patients in good and bad responders and to detect specific periods of the psychotherapy that could be of particular very good or very bad relationship quality. Finally, 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 versus 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. We can begin to draw some preliminary conclusions about using synchrony in order to classify the relationship quality of these videos.

### Conclusion

This work led to the development of R scripts to manipulate motion history, filter it ; adaptation of Python module of SyncPy. Too our knowledge, these are the first scripts that are published in open access to enable easy replication and further developments. We draw some recommendation to get psychotherapy material of better quality in terms of video and sound recording and psychotherapy techniques and outcomes. Furthermore, we suggest some modules that could be

added to extract more social signals from other modalities and precisions. With our method, we can began to draw some preliminary conclusions about using synchrony in order to classify the relationship quality of these videos.

See the full [pre-registration](#).

## Poster

# Extracting automatically social signals from psychotherapy sessions

Thomas Gargot<sup>1,2,\*</sup>, Giovana Varni<sup>2</sup>, Jonathan Aigrain<sup>2</sup>,  
 Mohamed Chetouani<sup>2</sup>, Michel Spodenkiewicz<sup>1,2</sup>, David Cohen<sup>1,2</sup>  
 1. Hôpital de la Pitié Salpêtrière, 2. Institut des Systèmes Intelligents et de la Robotique, Paris, France  
 \* thomas\_gargot@hotmail.com

## Introduction and motivation

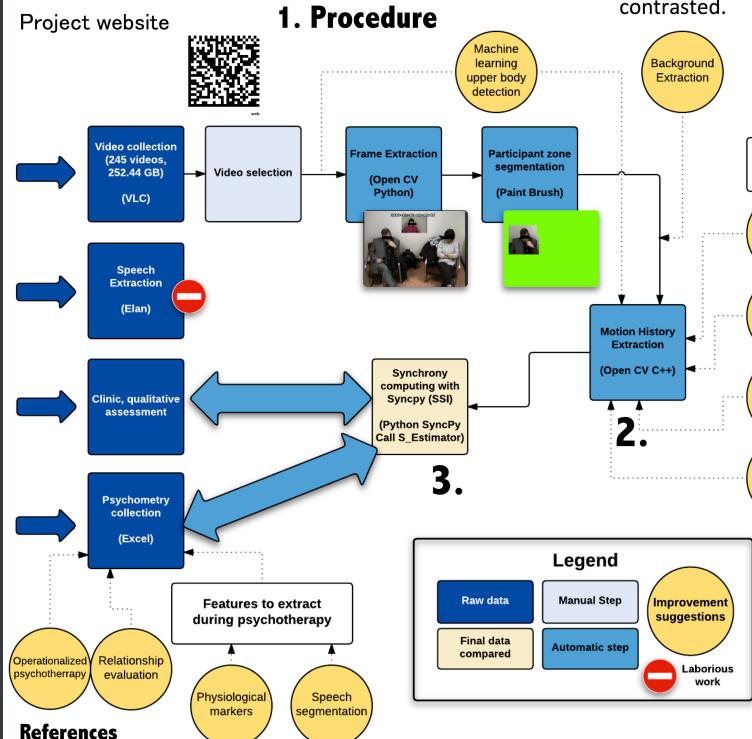
Psychotherapy is an important part of treatment of mental disorders alone or complementary with pharmacotherapy. Some techniques are now widely evidence-based and very cost effective (Layard & Clark, 2014).

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.

Project website



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).

## 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 developed 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.).

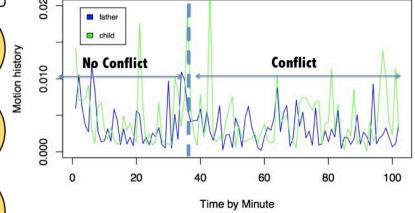
Using an interdisciplinary and open science approach, we developed some modules toward this goal that could be freely and easily re-used by other teams in other databases or with other modules.

## Future work

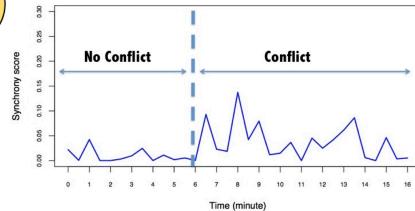
The next step would be to design a specific database anticipating some technical problems (speech segmentation, overlapping of subjects, micro-movements) and define more precisely different conditions of psychotherapy that could be contrasted.

## Example on a conflict video

### 2. Motion history



### 3. Synchrony score



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 SyncrONY 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 different configuration and theoretical frameworks (S Estimator technique). The package used for synchrony was developed by the engineers of the team and it was the first time that it was used in real world data. The input of psychiatrists and psychotherapists 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	Phd Student in social signal processing. He did scripts to extract frames from videos, to extract Motion History from videos frame by frame, knowledges in Python (OpenCV)
Nicolas Bodeau	Bio-statistician and informatician in ISIR and La Pitié, he supervised me for R and statistical analysis
Mohammed Chetouani	Prof in social signal processing, supervised the project, and the SyncPy library
David Cohen	Prof of Psychiatry, original idea, supervised the project
Catherine Saint Georges	Psychiatrist, knowledge in interpreting synchrony scores in psychiatry, helped with first speech annotation with Elan
David Reversat	Computer engineer, installation of SyncPy and associated modules, debugged SyncPy
Michel Spodenkiewicz	Psychiatrist, collected the databases from other teams (videos from INCANT and Monrado studies), direct supervision
Giovanna Varni	Post doc researcher Social signal processing, Development of the SyncPy library, supervised me for Python

The Monrado (Dr Coady Vulliez, Mrs Monika Szymanska) and Incant (Dr Phan, Mrs Bastard) teams recruited subjects and psychometric data.

Personnaly, I helped to define and operationalize the scientific question from the general idea, the tools and databases available. I did the bibliography based on the work of the team and Ramseyer team articles. I cleaned and prepared the databases. I adapted the Python and R script to make the extraction of the signal and performed the analysis. I did the figures and diagrams to explain the

process of the method and describe the results.

## Theoretical background

### 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 (hypnosis, psychoanalytical, cognitive-behaviorist, systemic ...) with very different histories, practices and level of evaluation. Contrary to other parts of medicine, psychotherapy and consequently psychiatry remains organized as schools which hardly interact each others. A lot of choices are done relative to the orientation of an hospital or a ward and not based on the best evidence based information which is often very limited in this field. Consequently, there isn't any consensual framework between psychotherapies.

Due to its history of being based upon scientific bases and in a scientific process of operationalize precisely the targets, Cognitive and Behavioral Therapies (CBT) are the most evidence based psychotherapy (Canceli et al., 2004 ; Layard & Clark, 2014) which is caused in great part by a lack of experimental work in other schools of psychotherapy.

Paradoxically, psychotherapies methods seems to be the most efficient treatment in medicine with a very low number needed to treat, i.e. it is necessary to treat only a few patient (around 2) to have a remission impact on a group of patient (Wampold cited by Briffault, 2009 ; Layard & Clark, 2014).

This could be striking to say that we are using useful and efficient treatments without knowing how they work. It is not an exception in medicine. Contrary to the classical biological and pathophysiological point of view, with Evidence Based Medicine (EBM), one don't need to know the mechanism of action of a treatment to prove that it is efficient. We only need to use experimental method (EBM approach) to compare it with an other treatment like a classical treatment or a placebo. For instance, one of the most widely used treatment paracetamol is a widely used analgesic and the proof of its efficacy is widely done. Unfortunately, in 2016, we can't say exactly how it is working on the body (Bonnefont, Courade, Alloui, & Eschalier, 2002).

Unfortunately, good psychotherapy is costly because it takes times and well trained therapist. Contrary to pills, psychotherapy isn't reimbursed in many countries. Public services hardly offer best practice of psychotherapy because of a lack of training and time.

Particularly, it is difficult to operationalize precisely what a psychotherapist does during a psychotherapy. Even the more operationalized psychotherapy as relaxation techniques, CBT deals with the particularity of each patient, its motivation, its context and don't have a priori a very precise roadmap. Despite these theoretical and evidence debates, there is almost a consensus that the relationship between the patient and the therapist. Establishing this therapeutic alliance is a very important part of the efficacy of psychotherapy (Horvath & Dianne, 1991 ; Martin, Garske, & Katherine, 2000 ; Cungi & Cottraux, 2006 ; Tschacher, Haken, & Kyselo, 2015). Even it is difficult to operationalize, some psychometric tools were developed like the Working Alliance Inventory (WAI) questionnaire (Corbière, Bisson, Lauzon, & Ricard, 2006). This led to the Dodo theory. It states, like

the judgment made by the Dodo bird after a race in the Alice and the Wonderland book, that all participants are equals. It would mean than nothing technical could differentiate a psychotherapy versus an other one, only the patient-therapist relationship would be important. This consensus is challenged by some behaviorist (CBT) that state that this determinant is only important when psychotherapist are not well trained and in poorly designed studies (Layard & Clark, 2014).

Therapeutic alliance is a first step before any other specific technique (Cungi & Cottraux, 2006). This process was a lot studied by the first psychotherapies like hypnosis. In a good therapeutic alliance, the contact is warm, authentic, collaborative. There is a lot of empathy, ie the therapist understand the point of view and the representation of his patient, shares it (contrary to apathy) while keeping in mind that he is not in the same situation that his patient and without being struggled by too much emotions (contrary to sympathy). One of the best operationalized and Evidence based kind of psychotherapy is motivational interviewing (Miller & Rollnick, 2013) which if much used in first steps CBT or psychoeducation. It can be the essential part of a work with addict patient or patients with addictions problems. The goal of motivational interviewing is to increase change discourse by stimulating it by some techniques:

- open question (like how are you today ?) with let the patient a wide range of possible answers contrary to close questions, specifically with Yes/no answers (like did you slept well yesterday night ?)
- Rephrasing what the patient is saying (eg If I understand well, you told me that..., isn't it ?) especially the change discourse.
- Summing up, the discourse of change of the patient (eg you told me that your consumption of cannabis wad a problem because your motivation was much lower)
- Avoiding direct advices but stimulating it by questions
- Accept the ambivalence in a change, the patient need to resolve it on his own without being pressured by the therapist
- Avoid judgement
- Searching for small and positive steps toward change
- Reinforcing any positive step
- Defining and raising the values of the patient and not the therapist values
- Giving advices only when it is requested by the patient

We can see that even this very operationalized psychotherapy let place for a lot of improvisation and adaptation to a precise situation and can not be standardized. Consequently, it needs a lot of practice and training beyond the theoretical knowledge. A lot of reflexes (like giving advices) need to be unlearned, that takes time. However, it is quite difficult to give precise feedback (supervision) without being involved directly in the therapy with the therapist. In complex settings like familial psychotherapies, the therapist can be directly supervised by a colleague behind a two-way mirror.

One way to improve learning process and evaluation of psychotherapy in practice is to make annotation of them. After filming psychotherapy, it is possible to evaluate what is exactly happening during this process. Some evaluation tools were developed like the MITI (Moyers, Manuel, & Ernst, 2014) for the motivational interviewing. Software like Elan can be very useful for that and enable to annotate several modalities at the same time. However, it can be very difficult and tedious.

In many parts of medicine, we don't directly evaluate the desired outcome of a therapy (like the mortality) because it happens too late. Consequently, intermediate marker help the physician to choose the best treatment. For instance, we can evaluate the efficacy of an anti diabetic on blood sugar mean reflected by Hb1Ac. This avoid to wait several years to check if the treatment is efficient or not (which can be evaluated only by disability created by diabetes or mortality). Consequently, it is very convenient. If we can get a automatic signal of this therapeutic alliance, we could get an intermediate marker of the efficacy of the psychotherapy which unfortunately doesn't exist at the moment. ~~However, we need to keep in mind that this intermediate markers are only intermediate and are not always relevant. Some anti-diabetics decrease Hb1Ac and increase mortality whereas classical one decrease both.~~

Fortunately, it is possible to evaluate the efficacy of a psychotherapy via two direct means. The first one are the direct outcomes of psychotherapy with behavior or consumptions (like number of cannabis joint consumed or number of panic attacks). This kind of easily reported outcomes are more the exception than the norm. However, it is not possible in all mental illnesses. Some psychometric data like Beck Depression Inventory (BDI) (Beck, Steer, & Carbin, 1988) for depression, or State-trait anxiety inventory (STAI) (Spielberger, 1970) were developed to evaluate the state of different pathological characteristics. These scales can be filled by the therapist or the patient directly. However, it is quite time consuming and not much done in practice outside research or CBT training.

To sum up, we can say that there are a lot of different kinds of psychotherapies. Some of them are largely evidence based and operationalized which can help for research. It is difficult and costly to train psychotherapist. Psychotherapies are efficient but it is hard to say exactly why. There are ways to evaluate its efficacy but they are not used since they are time consuming. There aren't many easy and quick ways to evaluate them.

## Potential role of cognitive sciences in improvement of psychotherapies

### Learning theory

Learning theories were developed by Skinner and Pavlov. They state that during a learning period, there are association between different stimulus. For instance, a behavior that will be positively reinforced will be repeated. A situation that is painful will be avoided. They are widely used in most recent cognitive studies in the term reinforcement learning. They are very used in cognitives sciences especially in modeling decisions process and understand brain activity. In psychotherapy, they are widely used in CBT approach to understand maintenance of a pathological process. Often a patient is being reinforced by avoiding a short term pain but this leads to avoidance of the situation and keep abnormal and painful emotions and cognitives schemas. The role of the therapist if to help the patient to became of this process during the functional analysis and to suggest him to change to healthier behavior by using new reinforcers.

### Cognitive neuroscience and education

With the progress of neuroscience there is a growing evidence toward some techniques to improve curiosity, attention, motivation of students (Dehaene, 2014). A new field of research neuroeducation is developing, a lot of these work could be used in psychotherapy too.

### Cognitive dissonance and engagement theory

Even if it wasn't described by the creators of the motivational interviewing, we can see that this process is very close to the engagement theory, and engaging communication developed by Beauvois and Joule (Joule & Beauvois, 2010) after the work on cognitive dissonance by Festinger (Festinger & Carlsmith, 1959). This theory was developed and validated in control people without any mental illnesses but could be very interesting in psychiatric patients and in Psychiatry settings (Fond et al, 2014). We can see consequently that we could use some very defined experimental settings like in cognitive and social psychology to improve our understanding of the process of psychotherapy. This is what will be used in the Monrado study to compare two situation without any conflict and with a conflict.

Cognitive dissonance theory is a counter-intuitive theory but very developed and replicated that deals with how we react to cognitive conflict and how we often rationalize our behavior in several situations.

Festinger asked participants to make a dull task. When the participant is freely making it for free, he feels a conflict between thoughts like "I am a clever guy free to make something" and its behavior "I am making something completely dull and useless". This conflict is difficult to handle. The subject is in a situation of cognitive dissonance. In this situation, he resolve his dissonance by changing his thoughts "this task is not the dull but rather interesting". However, in exactly the same situation, when the participants is paid, the effect disappear, there isn't anymore conflict. The subject still find the task completely dull as it was organized (Festinger & Carlsmith, 1959). After this classical paradigm, several explanations were developed especially by Beauvois, Joule who evaluated some other paradigms consistent with this theory (Beauvois & Joule, 1987, 1996, 2010). The Foot in the door technique (how to change a behavioral by very small steps), the "You are free" paradigm, the "labelling technique" (it seems your health is very important for you), "touch technique" were widely evaluated without much publicity inside psychotherapy field even if a lot of their approach can be consistent with Motivational Interview. Consequently, a behavior can be much more explained by a context and by the previous behaviors of the subject than on the expected outcome as expected by the reinforcement learning theory.

### Bandura

Bandura challenged the reinforcement theory too by demonstrating that imitation was a major process of learning. In the Dodo doll experiment (Bandura et al, 1963), he showed that the behaviors of children could be predicated by the models of adults they saw just before. This process is widely used in psychotherapy where the therapist can play the role of model in role play or exposure for instance which the patient can imitate.

### Mirror neurons, mirroring and empathy

Mirror neurons were discovered by Rizzolatti (Rizzolatti & Craighero, 2004). When a monkey seize an object, specific neurons of this task (in this receptor field) fire. The same neurons fire when this monkey see an other one making the same action. Consequently, it seems that there is a system in which the monkey simulate the outside world. This system is still activated when he really perform the task.

Neal and Chartrand showed that the mimicry of facial expression could be very useful in the process of empathy. They blocked expression of emotion with injection of Botox in subjects faces. In this condition, emotion perceptions of the subjects were impaired (Neal & Chartrand, 2011).

Consequently, we can say the perception can't be separated with action but both of them interact a lot in a very strong loop.

Beyond that, some researchers like Iacoboni stated that the system of minor neurons could be involved in mirroring and imitation, a system that could be involved in empathy (Iacoboni, 2009).

However and This role of mirror neurons is not consensual.

## Theoretical advantages and pitfalls of synchrony studies

"Synchrony refers to individuals' temporal coordination during social interactions" (Delaerche et al, 2012). However, definition of synchronies are very numerous. There is a possibility to switch from the social world to a mathematical definition but since the point of view is different, it doesn't always refer to the same thing. Interdependence of dyadic partners' behaviors was described by many terms in the literature (mimicry, social resonance, coordination, synchrony, attunement, chameleon effect, etc.) (Delaerche et al, 2012).

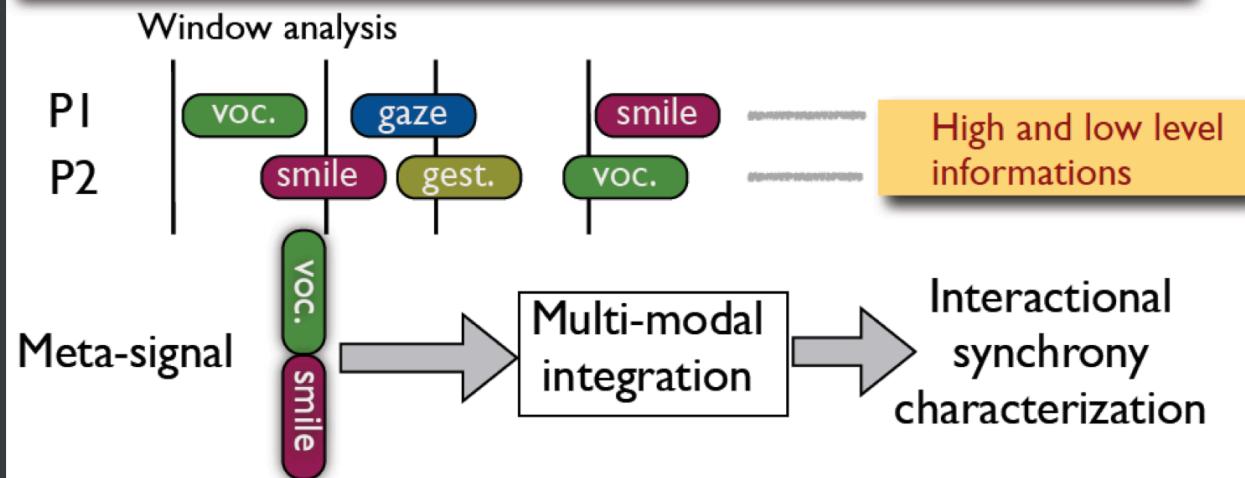
TODD Behavior matching [21]; mirroring; mimicry [22], [23], [24]; congruence and the chameleon effect [25] are related to convergence

Synchrony is related to the adaptation of one individual to the rhythms and movements of the interaction partner. For Delaherche, "synchrony is the dynamic and reciprocal adaptation of the temporal structure of behaviors between interactive partners" (Delaerche et al, 2012). It can be unimodal (eg. speech-speech) or multimodal (eg. speech-motor, someone answering to a sentence by a sign of the head).

Synchrony could be an important element of therapeutic relationship. This could be measured in any kind of psychotherapy without theoretical a priori. It is a social signal but brain correlates were found (Dumas, Nadel, Soussignan, Martinerie, & Garner, 2010) and hormonal correlates too (Weisman et al., 2013).

## Modeling interactional synchrony

- ▶ Evaluation of non-verbal behaviors produced by the partners
  - ▶ Integration of non-verbal behaviors produced by the partners



*From Human communication dynamics presentation, D. Cohen*

It was studied in psychotherapeutic settings by other teams. Ramseyer et al. investigated non-verbal synchrony between patient and therapist during psychotherapy sessions. They showed that non-verbal synchrony was associated with therapy outcome and patient's view of the therapy process (Ramseyer & Tschacher, 2011). They found that synchrony was increased in sessions that were considered by patient as sessions with high relationship quality and in patients experiencing high self-efficacy. Furthermore, higher non-verbal synchrony predicted psychotherapies with higher symptom reduction.

However, in these settings, there could be only one patient and one therapist. Furthermore, it is not possible to replicate directly this study since the technical aspects are not explicated and programs codes are not shared.

Some recordings are used in diagnostic phase where lack of coordination in non-verbal behaviors was found as a risk of recurrence of depression (Bouhuys & Sam, 2000).

Independently of the psychotherapy theory and framework, it seems that synchrony could be used in very different settings and could reflect the therapeutic alliance. In the future, we can imagine that synchrony score could be extracted automatically in different modalities (synchrony of speech, motion history, smiles, expressed emotions). That would give a feedback about the process of the psychotherapy that will be more objective and less time consuming than filling a psychometric questionnaire.

It was quite difficult at the beginning to replicate the studies that were done by other teams and to find these two databases that could be relevant for this goal because of quality of the recording and quality of psychometric data collected. After we had our own scripts and procedure for Incant database, it was very easy to replicate the analysis for Monrado database.

To avoid this lost of time, and to improve validity and replicability and improvements of our work. we tried to develop an open science approach. The synchrony computation software SyncPy developped by the team G. Varni and M. Avril was published on open source on GitHub. We decided to continue this approach to improve accessibility, transparency of this work to help future improvements.

The idea of this project is to be a first step toward a better understanding of therapeutic or parent-children relationship. We focused on motion history since it was technically the easiest modality. However, a lot of improvements could be added to add other modalities or to improve analysis by using for instance modeling or machine learning techniques. Therefore, we decided to share our analysis scripts. Both of them are published but the Monrado script is much more commented, organized. Data dictionnaries are more described in order to better understand the process.

However, it is not possible to share the raw materials with consist of confidential psychotherapy sessions. It would be very relevant to share this material to let possibility to other team to compare this results with other softwares or modules with different goals than motion history synchrony analysis. In the research in algorithm development, some databases are shared and promoted during challenge sessions between laboratories. This enable the development of the best classifier for instance.

## INCANT Study

### Introduction

Cannabis dependence psychotherapy is important because cannabis is the most used illicit drug worldwide (Danovitch & Gorelick, 2012). This substance leads to cognitive troubles (motivation, social withdrawal) that could be harmful for learning process or to find a job. Furthermore, it increases the risk of schizophrenia (Large, Sharma, Compton, Slade, & Nielssen, 2011).

### Cannabis addiction

Cannabis dependence is defined by DMS V (Diagnostic and Statistical Manual of Mental Disorders, (American Psychiatric Association, 2013) by:

"A. A problematic pattern of cannabis use leading to clinically significant impairment or distress, as manifested by at least two of the following, occurring within a 12-month period:

1. Cannabis is often taken in larger amounts or over a longer period than was intended.
2. There is a persistent desire or unsuccessful efforts to cut down or control cannabis use.
3. A great deal of time is spent in activities necessary to obtain cannabis, use cannabis, or recover from its effects.
4. Craving, or a strong desire or urge to use cannabis.

5. Recurrent cannabis use resulting in a failure to fulfill major role obligations at work, school, or home.
6. Continued cannabis use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of cannabis.
7. Important social, occupational, or recreational activities are given up or reduced because of cannabis use.
8. Recurrent cannabis use in situations in which it is physically hazardous.
9. Cannabis use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by cannabis.
10. Tolerance, as defined by either of the following:
  - a. A need for markedly increased amounts of cannabis to achieve intoxication or desired effect.
  - b. Markedly diminished effect with continued use of the same amount of cannabis.
11. Withdrawal, as manifested by either of the following:
  - a. The characteristic withdrawal syndrome for cannabis (refer to Criteria A and B of the criteria set for cannabis withdrawal, pp. 517-518).
  - b. Cannabis (or a closely related substance) is taken to relieve or avoid withdrawal symptoms.

[...]

Specify current severity:

305.20 (F12.10) Mild: Presence of 2-3 symptoms.

304.30 (F12.20) Moderate: Presence of 4-5 symptoms.

304.30 (F12.20) Severe: Presence of 6 or more symptoms." (American Psychiatric Association, 2013)

### Cannabis addiction treatments

There are several researches for pharmacological intervention (agonist substitution, antagonist and modulation of other neurotransmitters pathways) but none approved nor recommended.

Psychotherapy is the better treatment accessible in 2016. There are evidences for some of them:

- Motivational Enhancement Therapy (MET or Motivational interviewing MI) and Cognitive and Behavioural Therapies (CBT) in which I am trained
- Contingency Management (giving positive reinforcement to support change)
- Familial and systems interventions, in which Multidimensional family therapy (MDFT) which is very comprehensive and much evidence based. A 224 adolescent randomized trial showed that MDFT was more effective than CBT (Liddle, Dakof, Turner, Henderson, & Greenbaum, 2008) but a big study Cannabis Youth Treatment didn't. INCANT study is an international study evaluating its efficacy (Danovitch & Gorelick, 2012).
- In a Cochrane review, one of the most consensual Evidence Based Reviews, (Denis, Lavie, Fatseas, & Auriacombe, 2006) found 39 studies considered eligible and selected only 6 quality studies with enough quality on psychotherapeutic intervention for cannabis abuse and/or dependence in outpatient settings. Pooled together, these six studies represent 1297 patients. They confirmed the efficacy quality of CBT and MET. Contingency Management can be

combined with these therapies. Systemic and familial therapies studies weren't selected in this analysis.

## MDFT and the INCANT study

The MDFT is a "family based and developmentally oriented outpatient treatment for adolescent substance use disorders and associated or related problems". ([Incant.eu](#)) It was developed in Miami, US since 1985. It is an integrative group of interventions used in a lot of situations. In MDFT, sees the problem of cannabis consumption as a reflect of deleterious lifestyle. Consequently, the therapist targets as many life domains and problem behaviors as possible.

The INCANT study was a multi-site, trans-national randomized controlled trial led in five countries (Belgium, France, Germany, the Netherlands, and Switzerland) aiming to compare the efficacy of the MDFT versus Individual therapy. Randomization was stratified as to gender, age and level of cannabis consumption. After a pilot study and a pre-registration, it recruited 450 adolescents from 13 to 18 years old. It showed that MDFT was more efficient than Individual Therapy (Rigter et al., 2013). It is very rare to run such an international study with high level of scientific setting in psychotherapy.

## INCANT Study in France

We didn't collect the videos of all the centers, instead, we had only access to the videos from French centers. There were 2 centers with different configurations. Videos were recorded to check if the therapist was applying the psychotherapy that was assigned to each family. Technically, it was not supposed a priori that they would be used to evaluate social signal. For some subjects, there are a lot of videos but for others, only a few were recorded.

## Material and methods

### Videos

- We collected a 252.44Gb database of 277 Videos from 32 patients and their families, with a rate of 25 frames by second.
- They are encoded in VOB format.

### Softwares used

- Excel tables manager
- Powerpoint (figures)
- C++ Programming language used to extract motion history
- R Statistical software, filtering of the data frames, graphs
  - xlsx package
- Python Programming 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 computing synchrony scores
- FFMPEG Module for converting videos
- 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
- Lucidchart (figures)
- Typora : Markdown editor
- Git
  - GitHub website public version control repository

## Psychometric data

Due to technical problems we got the cannabis consumption data very late.

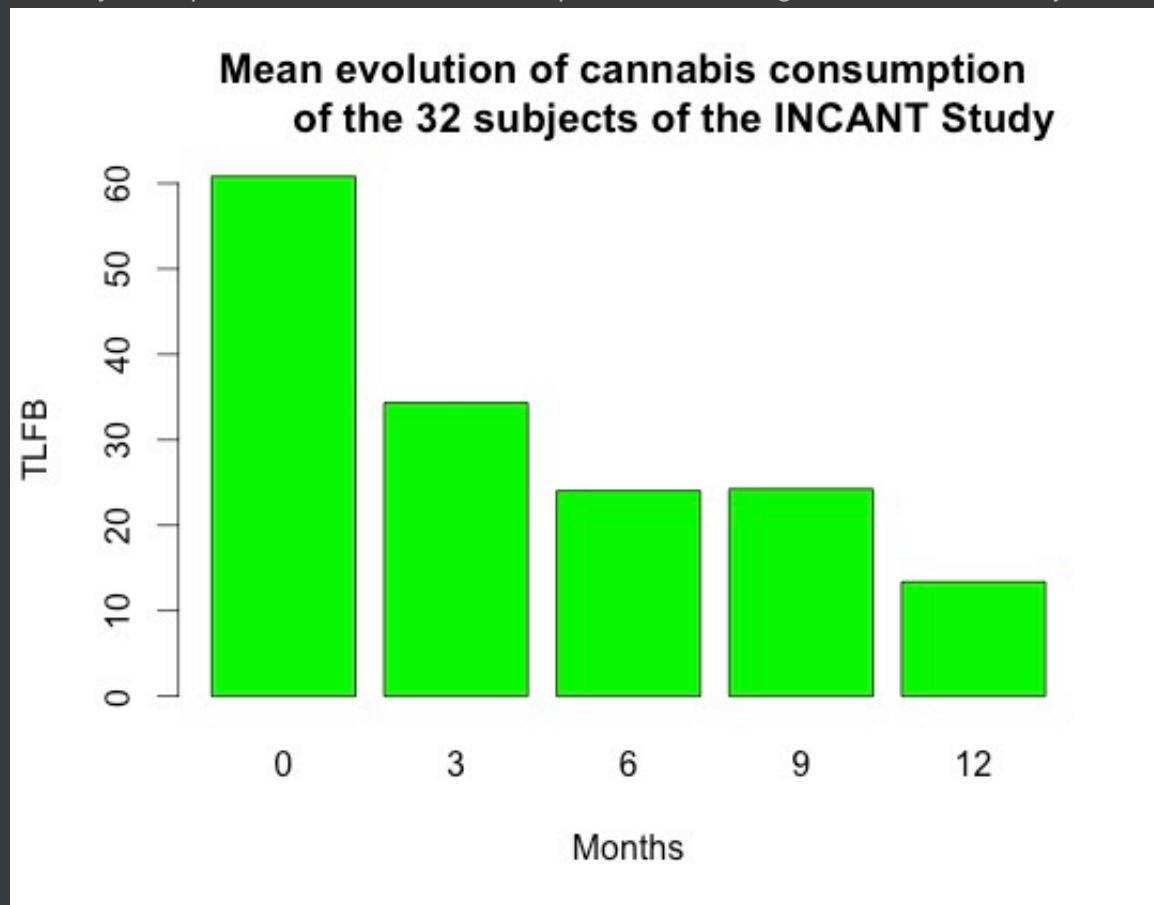
However, we didn't have any data dictionary with the definition of all variables and the exact questionnaires used to collect this information, the coding of Non Available data that may be -99. We planned to meet the team that wasn't unfortunately available at that moment. A lot of videos were lost or not recorded. We couldn't have access to the randomization to know which subjects received MDFT and which received Individual Therapy.

- These files 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).

We decided to focus on the primary outcomes of the psychotherapy the cannabis consumption, i.e. the TLFB.

These data were imported via the XLSToCSVConvertor.R R script which I wrote.  
It returns a dataINCANT.csv csv file and a dataCannabis.csv file.

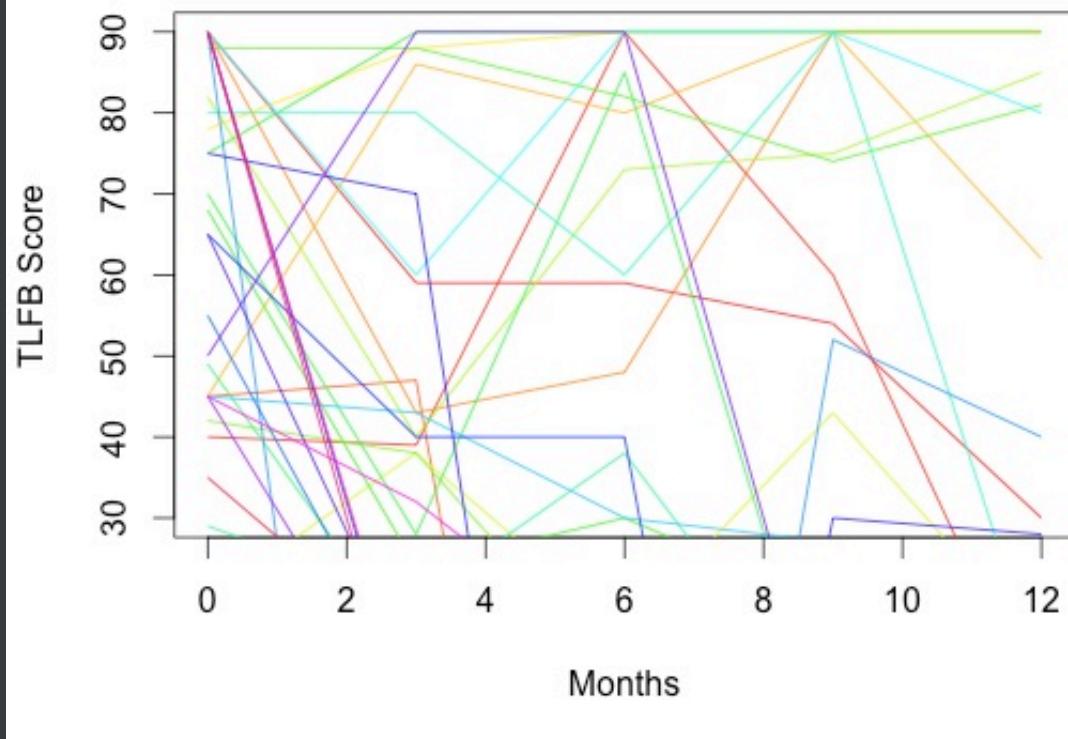
Globally, as expected, the cannabis consumption is decreasing with time for the subjects.



We tried to cluster 2 groups to contrast the synchrony scores between good responders and bad responders. Unfortunately, when we looked at the evolution of the cannabis consumption, it wasn't very easy to define good and bad responders.

<b>case</b>	<b>baseline</b>	<b>3.months</b>	<b>6.months</b>	<b>9.months</b>	<b>12.months</b>
F1002	90	59	59	54	30
F1007	40	39	90	60	5
F1009	45	47	-99	3	-99
F1014	90	43	48	90	90
F1019	45	86	80	90	62
F1023	78	88	90	90	90
F1027	20	38	16	43	14
F1044	82	40	73	75	85
F1069	42	38	9	21	20
F1070	88	88	82	74	81
F1073	68	24	30	19	11
F1085	75	90	90	90	90
F1098	70	28	85	0	-99
F1101	49	13	6	5	3
F1104	29	20	38	4	5
F1112	80	80	60	90	10
F1113	90	60	90	90	80
F1117	45	43	30	27	7
F1120	90	-99	-99	52	40
F2003	55	8	8	-99	3
F2005	65	40	40	-99	-99
F2010	90	3	14	4	10
F2011	75	70	-99	30	28
F2013	65	10	20	15	18
F2017	50	90	90	0	0
F2018	45	3	0	0	0
F2026	20	0	0	0	0
F2027	45	32	12	24	18
F2031	90	4	5	8	7
F2032	5	0	0	-99	-99
F2035	90	0	-99	0	0
F2039	35	13	0	15	17

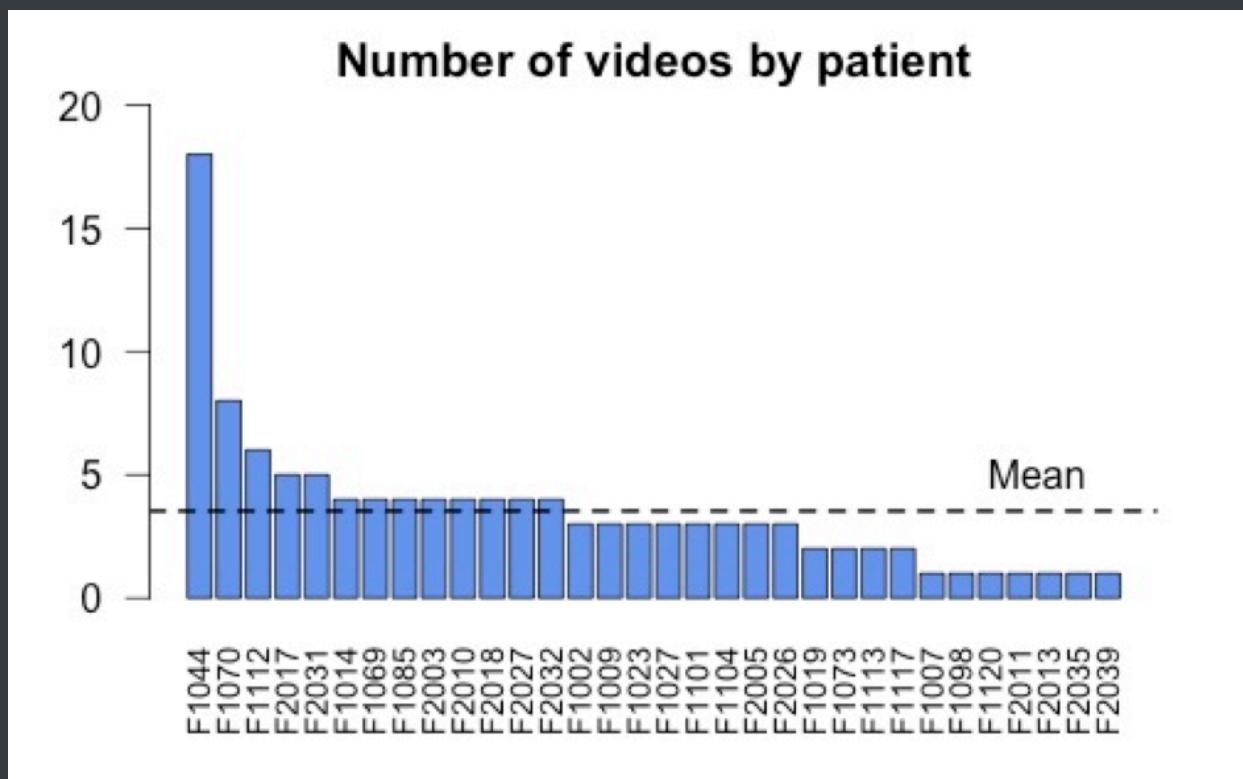
## **Evolution of cannabis consumption of the 32 subjects of the INCANT Study**



### Results

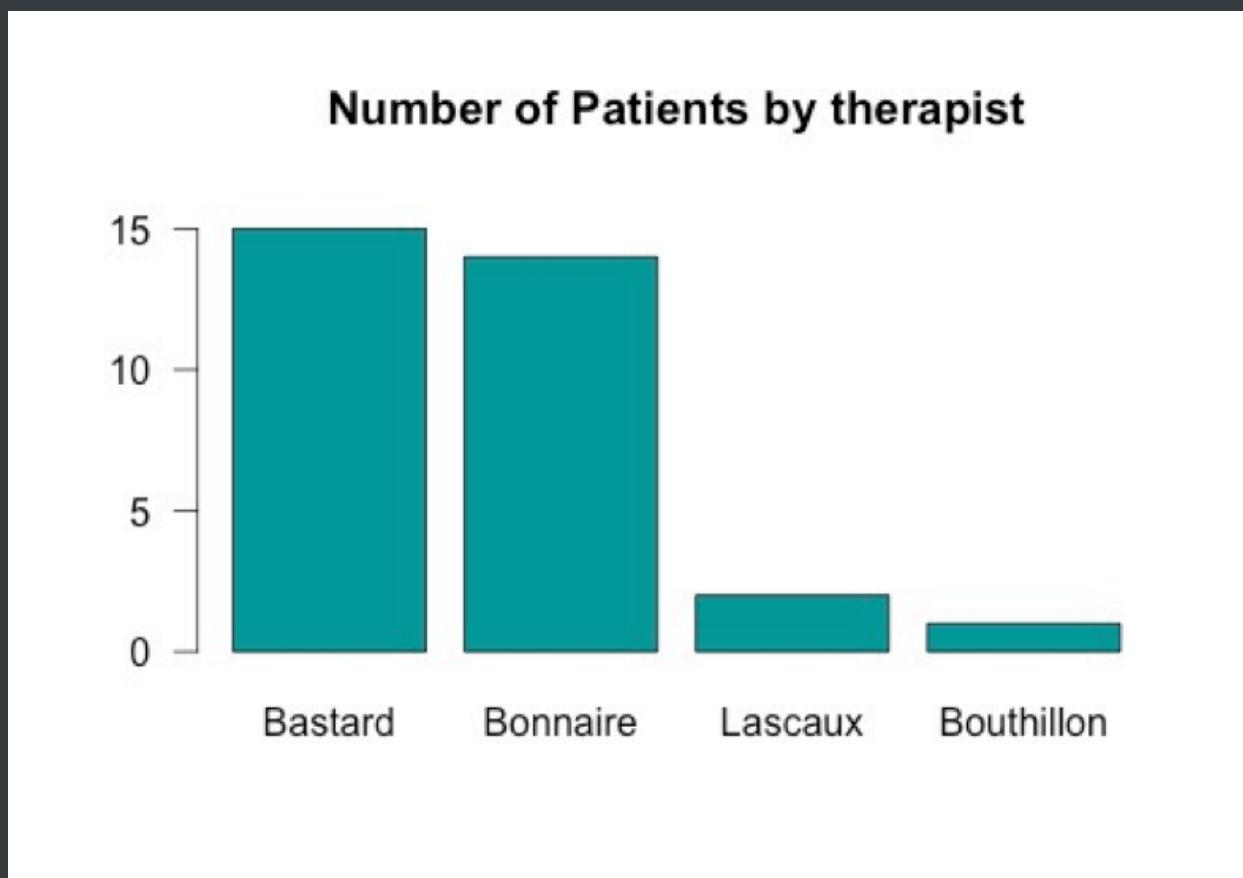
Pilot study: F1044 subject, his father, mother and therapist

We decided to make a pilot study on the most complete subject out of the 32 patients in order to evaluate the feasibility of the concept and to develop the process especially the R scripts. We chose the F1044 subject since we have the most videos (18 videos) with him or his family.



We can ask ourselves if this subject representative:

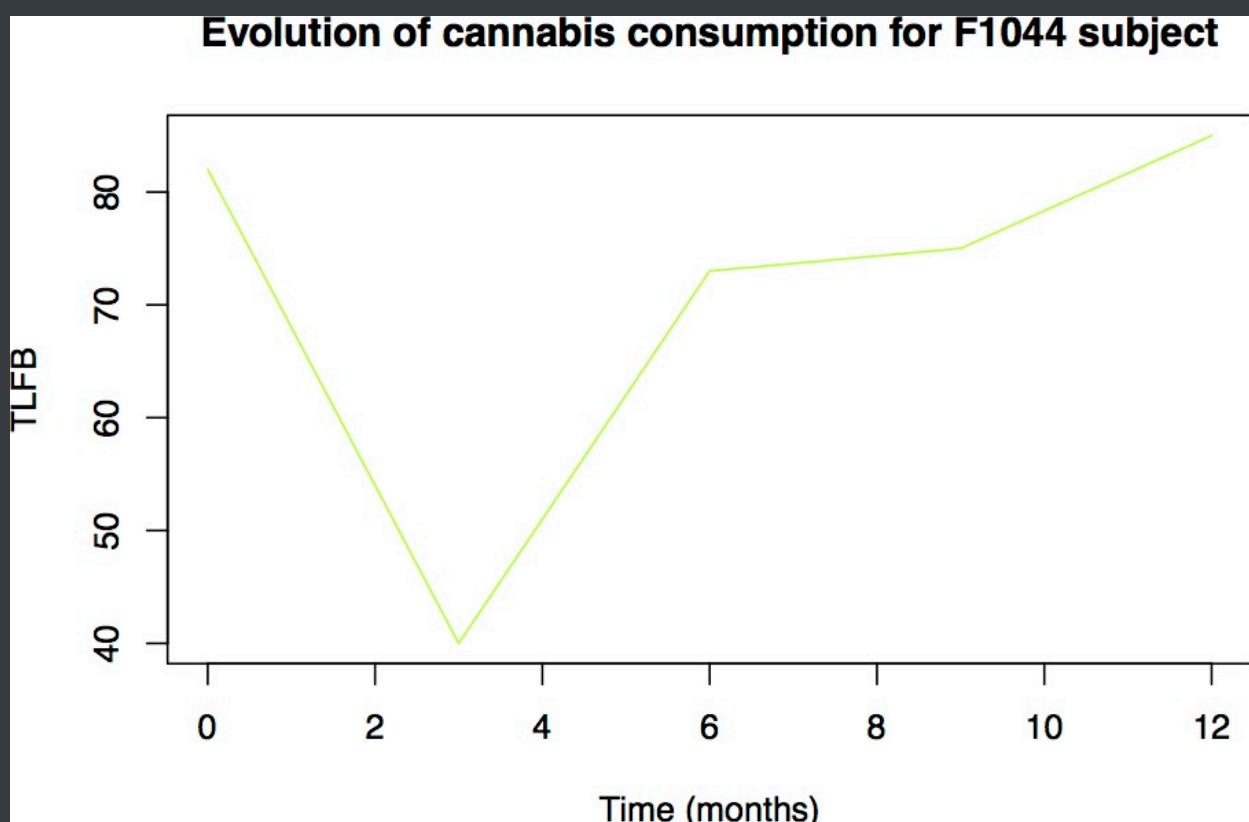
- The therapist is the first most representative concerning the number of patients and the first concerning the number of patient videos.



- 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).

Consequently, we can say that this subject is very representative of the sample. Unfortunately, the first, second, 10th and 11th videos were lost.

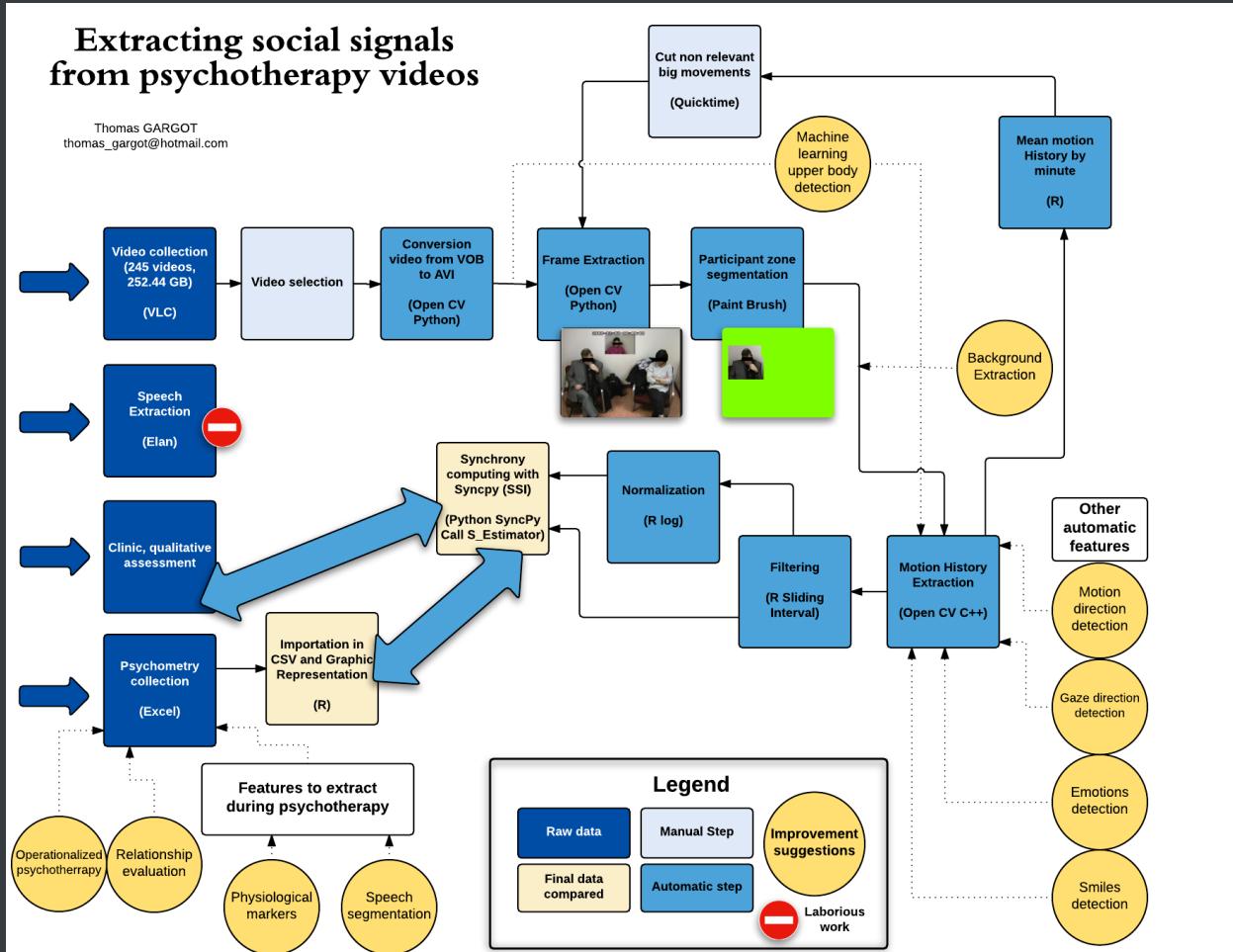
Furthermore, when we got the cannabis consumption evolution, we saw that the efficacy of the psychotherapy was not very clear and the evolution of it wasn't straightforward. Consequently, if our hypothesis of good synchrony or good evolution of synchrony as predictor of good quality of psychotherapy was good, we couldn't really expect any of these results in the F1044 subject when we see the objective evolution of the psychotherapy.



TODO que points, abscisses à points intéressants

#### Description of the database

Different steps of the social signal extraction



This figure summarize the databases we collected (dark blue), different steps that were done (blue and light blue) and some suggestions (yellow) that could enable to extract social signal from psychotherapy videos.

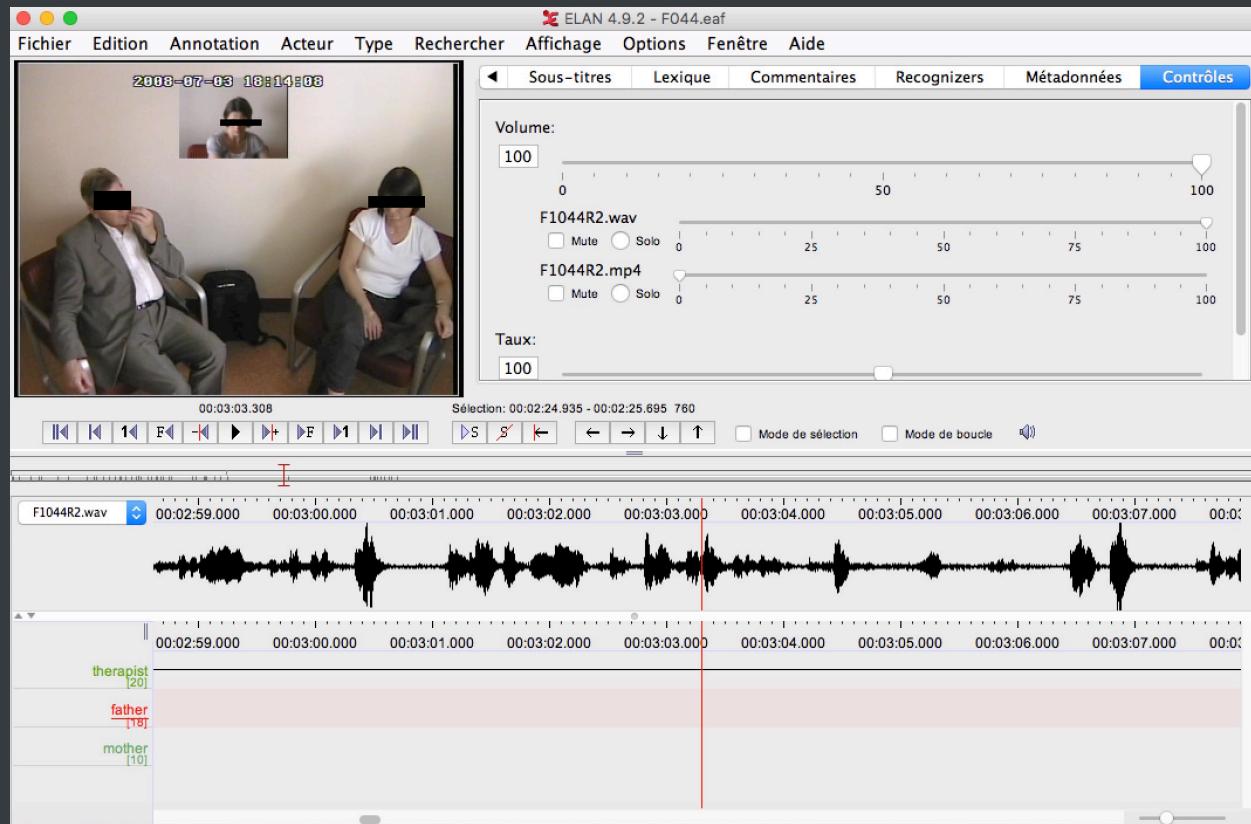
#### Verbal signal extraction to compute verbal synchrony

We began to extract social signal from videos with annotations made with the Elan software. The idea was to annotate the speaking of each participants. We used [Free Video Converter](#) to get a waveform and import the video and the waveform. This waveform helped me to identify the period of speaking versus no speaking unfortunately couldn't help to identify the source of the sound.

Most often locutors were filmed so it helped to detect manually who spoke but it wasn't always the case. Quite often, participant speech overlap that made the process quite difficult. There was some short onomatopoeia that were sometimes quite arbitrary to annotate or not. The end of the sentence was not always obvious with long vowels with not precise end especially during the overlap of speeches. It was made from the video and the waveform extracted from the video in an audio file.

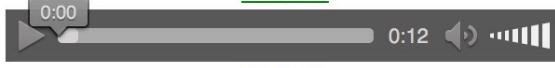
Then you can annotate when someone is speaking. This process is very fastidious (1 week work for one hour video at the beginning and so was gave up). Even when speaking with more trained annotators, the productivity is not much better.

For the moment, when we focus on this strategy, it is often necessary to extract manually some features even because of the lack of automatic process or in case of development, training and evaluation of these tools.



We tried to find ways to make source separation automatic even if it couldn't be perfect. We found the [FASST software](#). It is very powerful in order to separate sounds coming from different music instruments. However, even if the source separation of several music instruments is very efficient, differentiating automatically two voices at the moment is too difficult. We contacted the author of this software. He told us that this material is too difficult to separate at the moment since two voices are much more similar than two different music instruments.

## Source separation examples

<b>Linear instantaneous mixture (blind separation)</b>	<b>Professionally produced music recording (non-blind separation)</b>
	
<u>FASST Matlab script</u>	<u>FASST Matlab script</u>
<b>Separated sources</b>  <u>Source 1</u>   <u>Source 2</u>   <u>Source 3</u> 	<b>Separated sources</b>  <u>Bass</u>   <u>Drums</u>   <u>Melody</u>   <u>Remaining sounds</u> 

Example of source separation using FASST on a music sample.

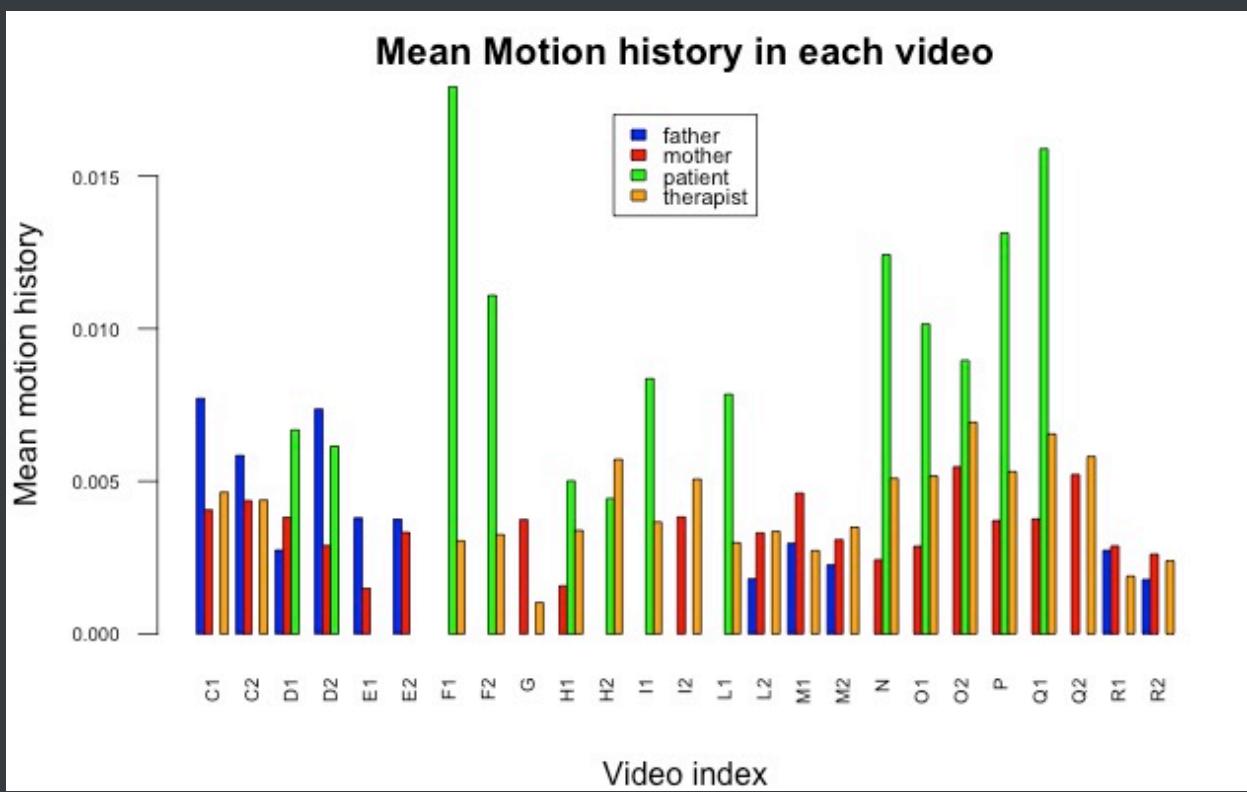
However, this separation problem can be anticipated during the recording of the videos. Several tie-mikes could be attached to each participant.

Consequently, we decided to give up this approach.

### Spatial organisation

In the first center, psychotherapy sessions consisted of filmed familial psychotherapies with 2 to 5 peoples organized in a circle. They are in samm room with a two-way mirror. Behind it, there is the supervisor of the psychotherapist that can call the therapist to give her some advice. The family is aware of this oragnization but this supervisor is not seen or known by them.

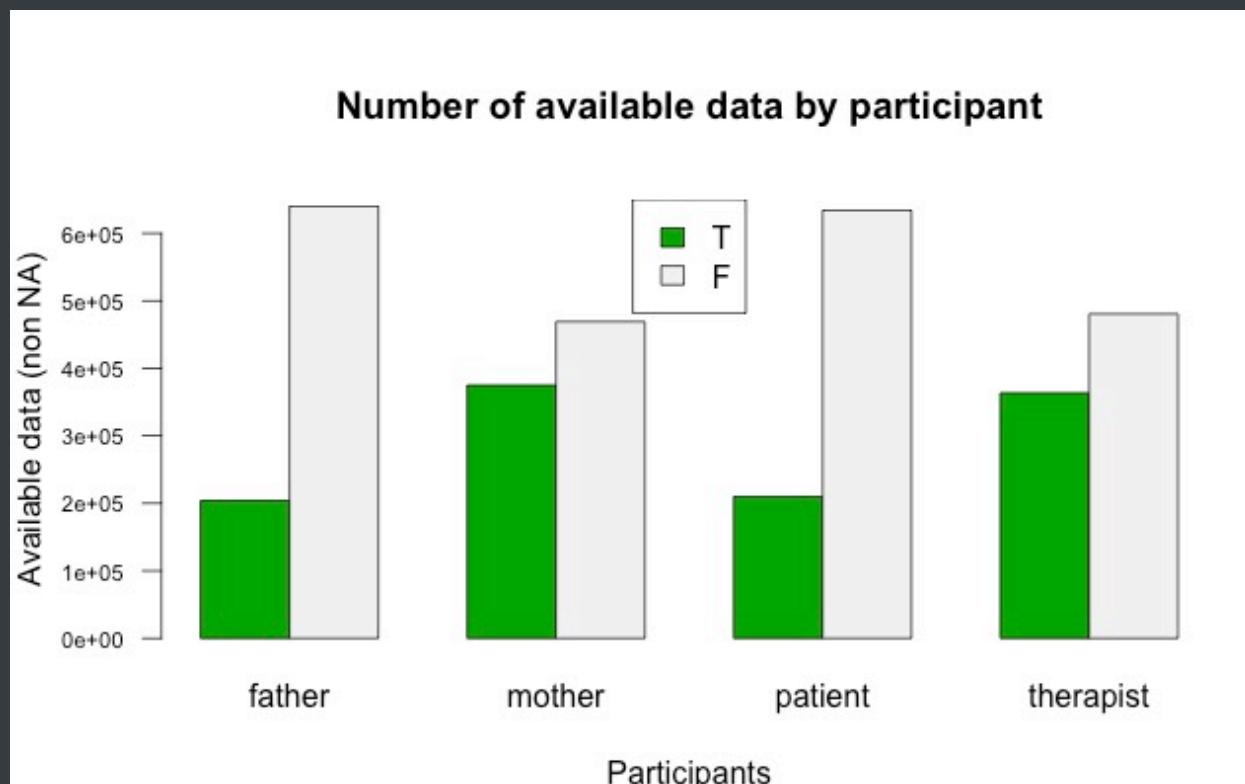
Two participants of the psychotherapy were filmed from the front. Another subject (often the therapist) is filmed from the other side. Its video is embedded in a window. We can notice that in this video, there isn't the patient but his two parents since he which wasn't motivated enough to come in this session, which is not so problematic from a theoretical point of view in systemic therapies. It is still possible to do a pschotherapy with this settings since we consider the communication and the attitude of each member of the family and we don't consider only the patient.



TODO : change to True/false plot

In this barplot of the mean motion history for each subject for each video, we can see the configurations of the participants involved for the F1044 subject for each video. These configurations can be very different. There are between 2 and 3 participants. All these configurations could change (eg patient-mother-therapist then mother-therapist) during the psychotherapy. Consequently, it was very difficult to compare the synchrony between different configurations. We can never get the information from the all 4 participants involved at the same time even if they are all involve like in D2 video since the psychotherapist wasn't filmed because of technical reasons. 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).

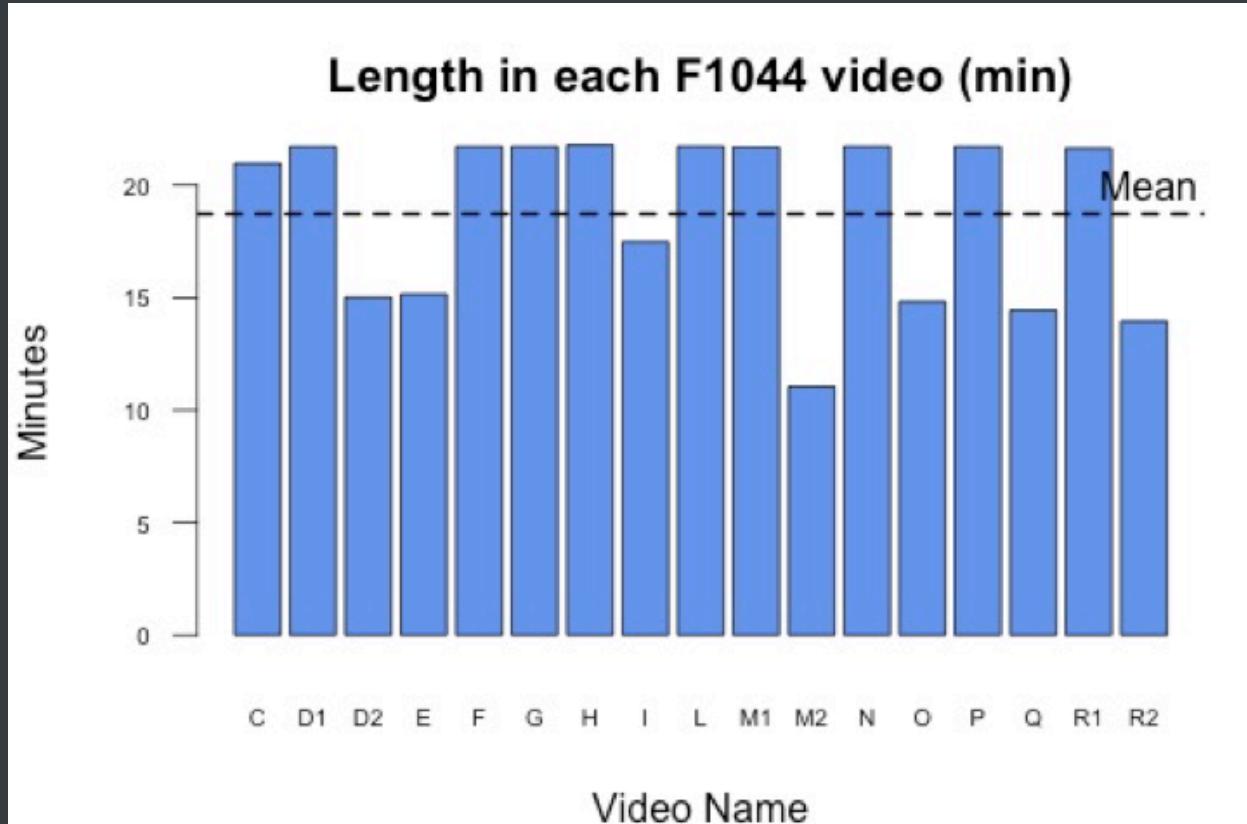
When we sum up the data we get, we can see that the mother and the therapist are the most present participants.



#### 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 during the psychotherapy to communicate between the main psychotherapist and the supervisor behing a two-way mirror and the therapy stopped at that moment. Sometimes, the participants made some pauses or the same video was used for different patient.



Raw data

#### 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 FFmpeg.

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

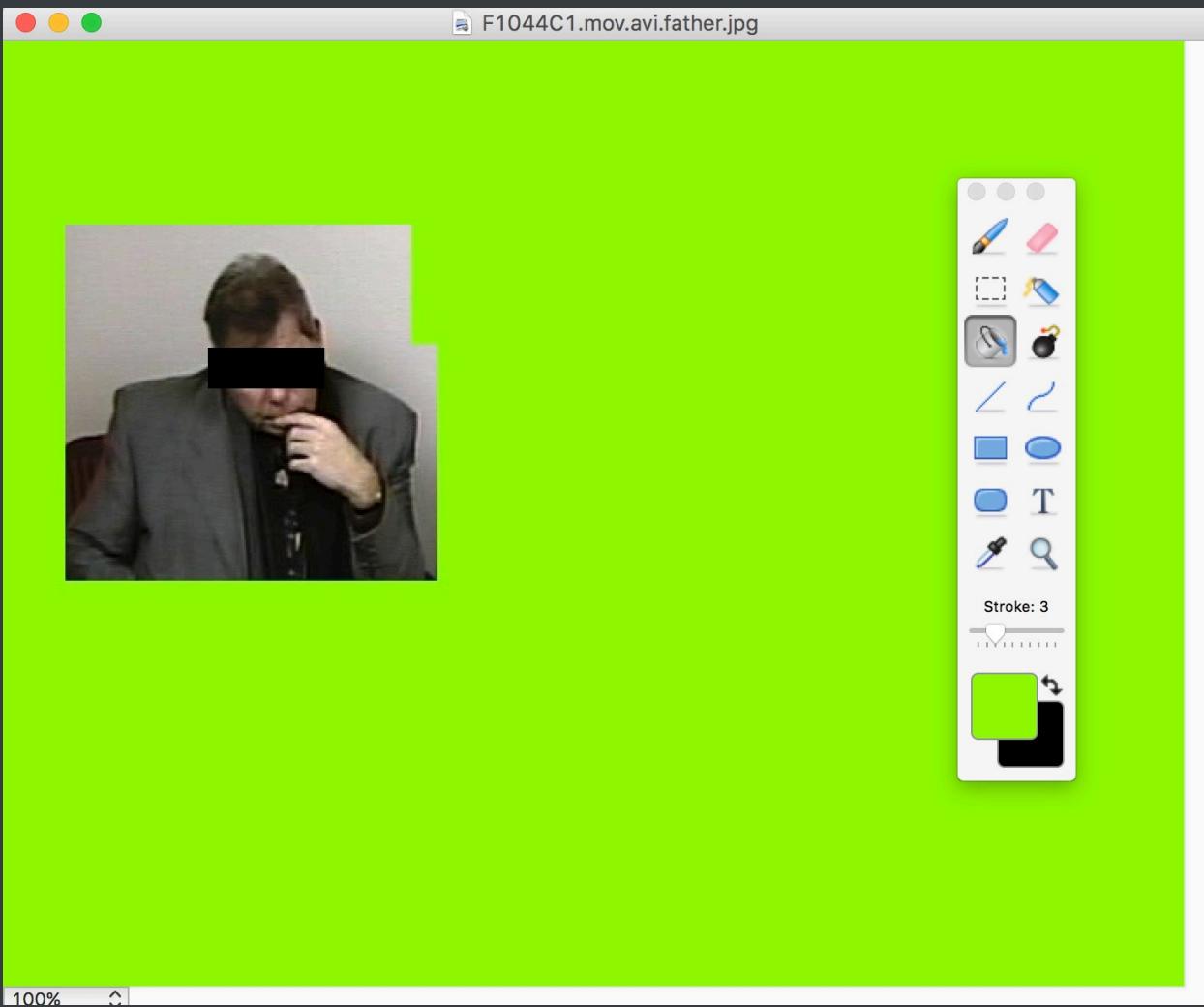


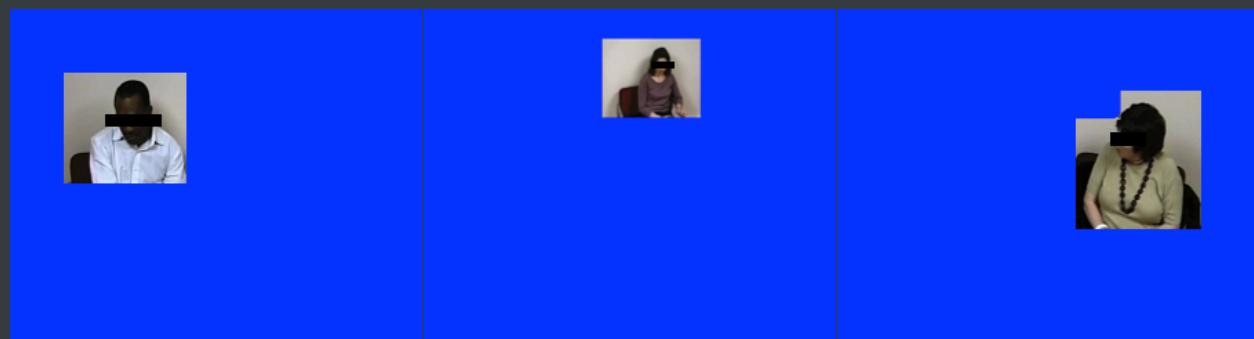
It was suggested to take a mean image of the video to superimpose the different configurations of the subjects on one video to be more precise during the spacial segmentation instead of taking a frame from the beginning of the video.

It would be much more relevant to extract the exact 3D position of the skeleton instead of the raw pixel. This can be extracted quite easily by a kinect. In a complex setting like this, it could be useful to use even several kinects for each side of the room or even better for each subject. A lot of other features could be extracted but the volume of this data can be very heavy with several megabytes by minute of video. This is much more difficult technically than raw videos alone. Some algorithms can extract relevant social signals from videos where they are recorded from the front (smiles, emotions, direction of the head (Hupont, Baldassarri, & Cerezo, 2012 ; Torre et al., 2015).

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

We selected a part of the video with each participant (father, mother, patient, therapist) with Paintbrush. We took only the upper part of the body of each subject to compare them since you can see only the upper part of the therapist. Consequently, we had one mask for each participant, so between 2 and 3 masks for each video.





### Technical problems

Unfortunately, since the videos weren't recorded for this specific purpose, we can notice several pitfalls.

The zoom level of the parents and the zoom level of the therapist was different. We'll see that they are hardly comparable.



In this frame, we can notice that there is an overlap between the region of interest of the therapist (in the upper part, 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 and the time with the patient frame (top).



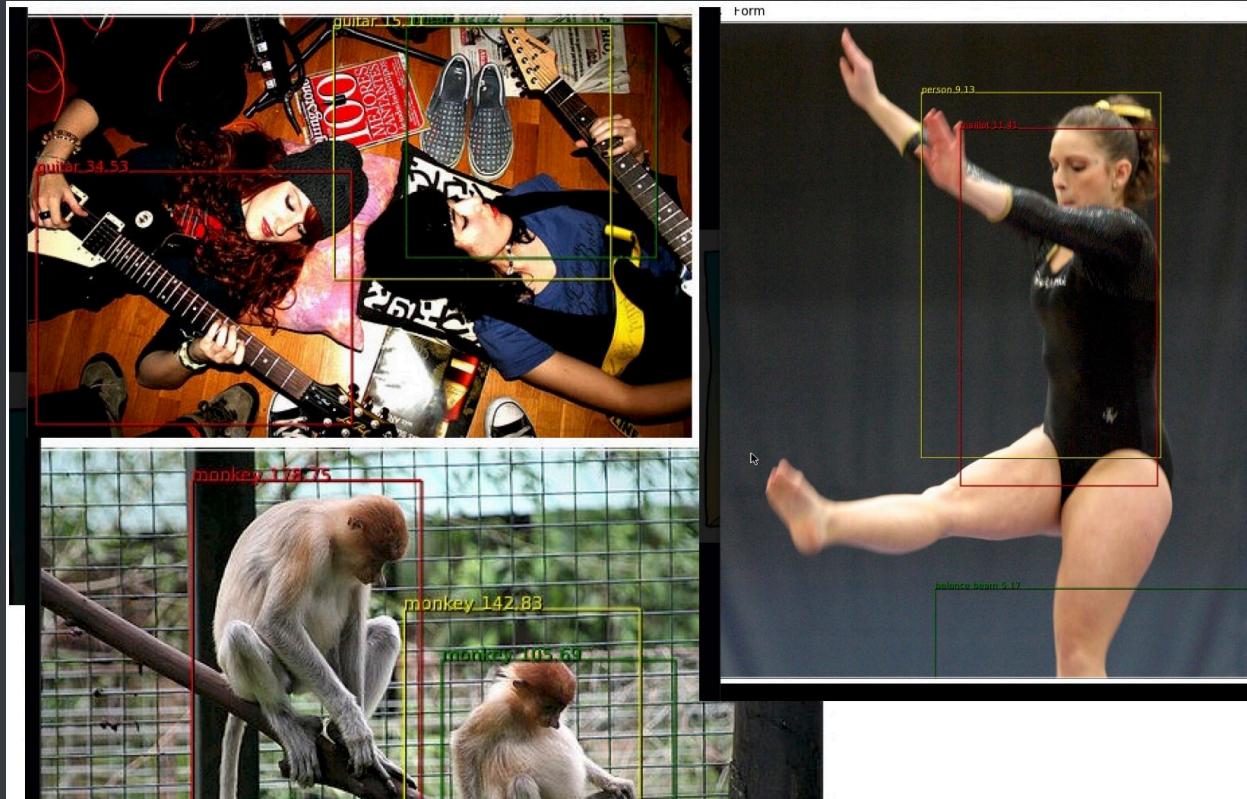
In this frame, we can see that the date is overlapping with the therapist. This was a big problem since the time was always changing and changed the pixels in this zone.

Sometimes, the camera is moving during the film. Obviously, this leads to a shift of pixel and consequently a signal of motion history even if nobody is moving. Consequently, it is very important to keep a motion less camera or cameras.

Sometimes, there are big movements of the subject which are not relevant for us. The therapist speaks with an other one via a phone, the participants left the room or

Participants can move during the video (change their seats, leave the room, even several families were recorded on the same video). 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 to do temporal segmentation/anotation.

In the future, it would be useful to use 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*.

With these masks and videos, we could extract motion history with a C++ script. It returns CSV file of the form "*F1044C.VOB\_res2.csv*". See files [here](#). We can notice that we can change the value of the median filter. In order to be able to compare several subjects, we decided to divide the motion history extracted by the size of the frame selected, it correct partly the zoom problem.

The motion history doesn't have any unit, it is the percentage of change of pixels change among all the pixels selected (non masked in green here).

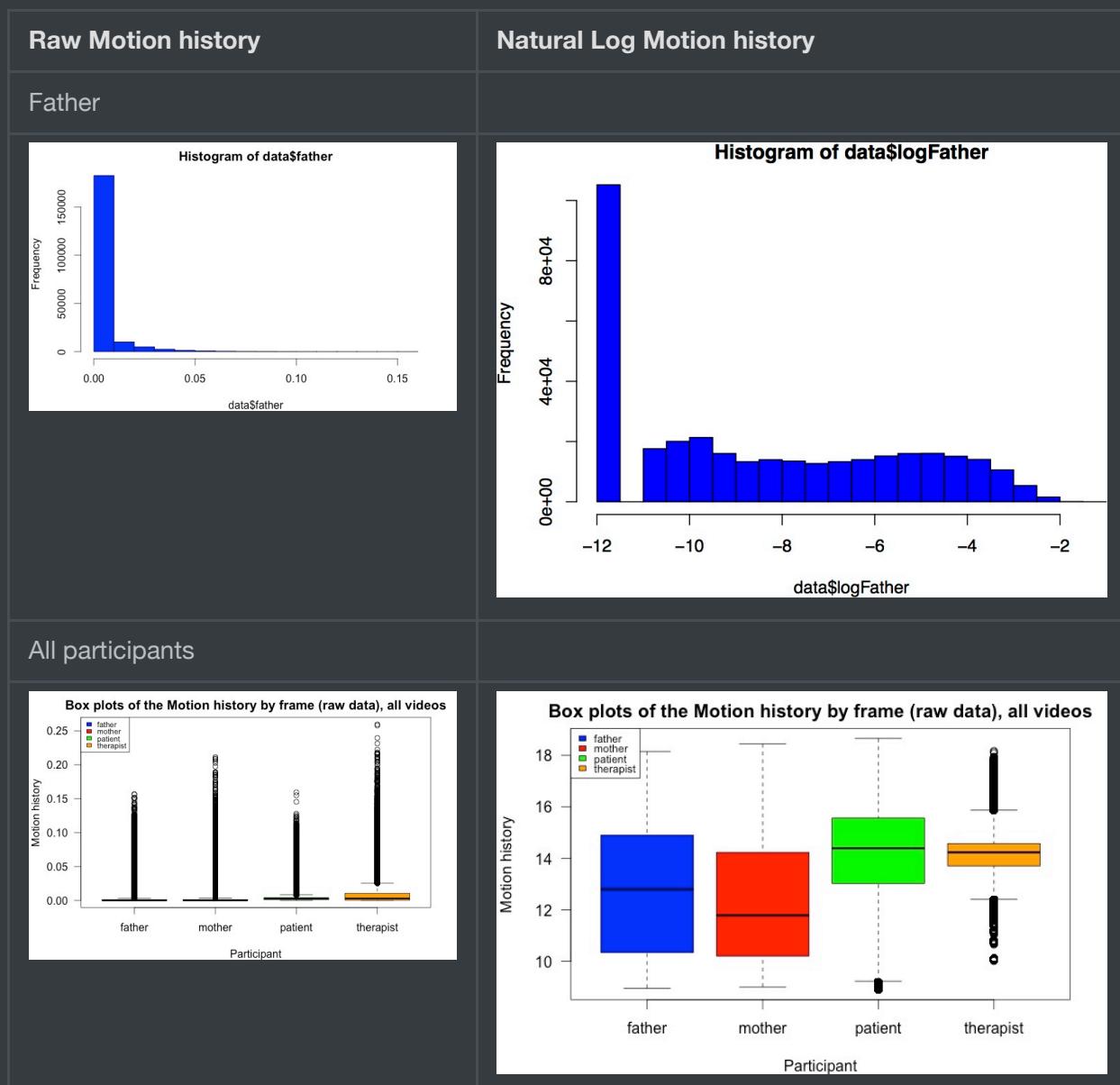
Headers :

- **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

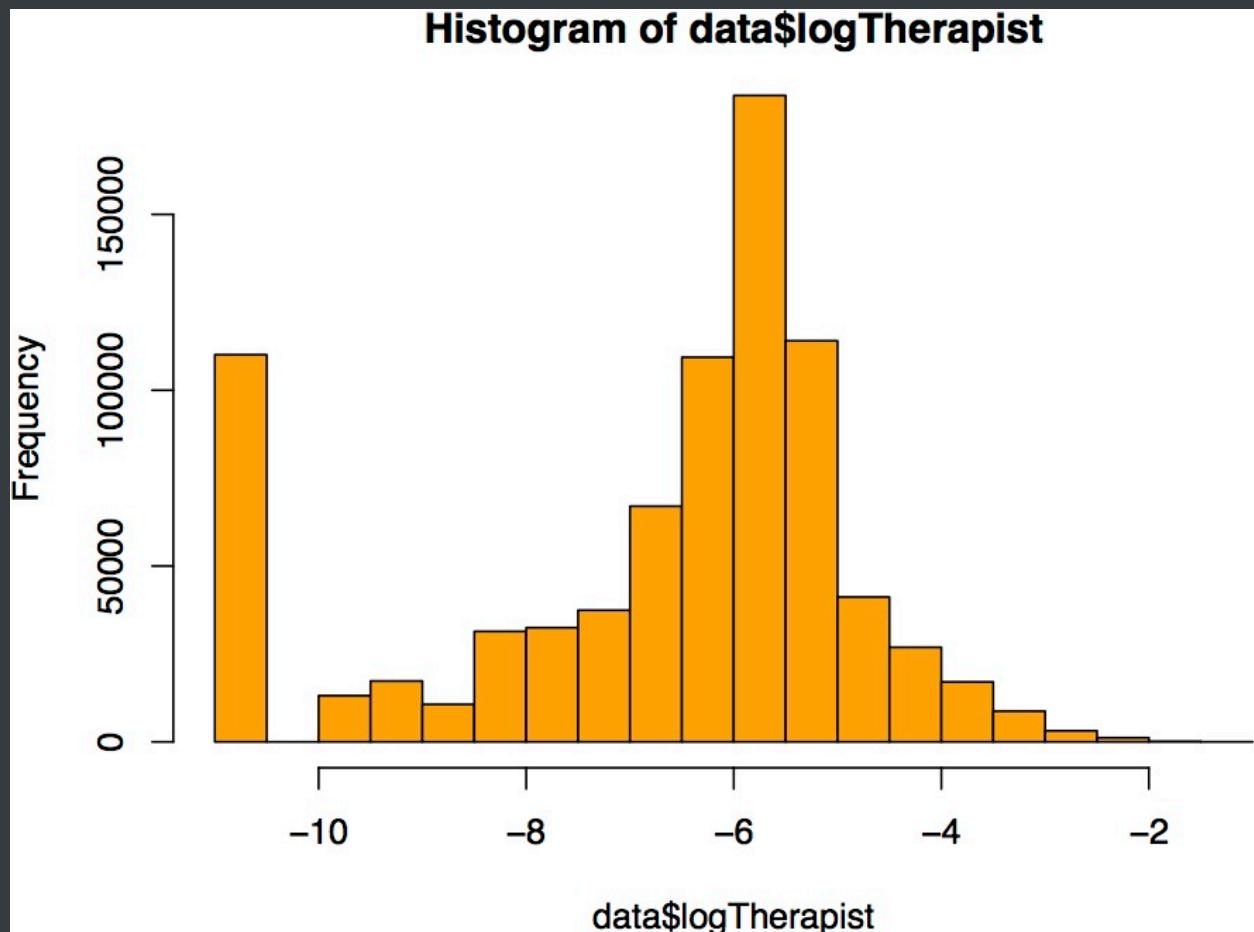
from quicktime when cut are made) and then to AVI, there is a problem of encoding and the length of the video is not well recognized and NA lines are generated that correspond to nothing.

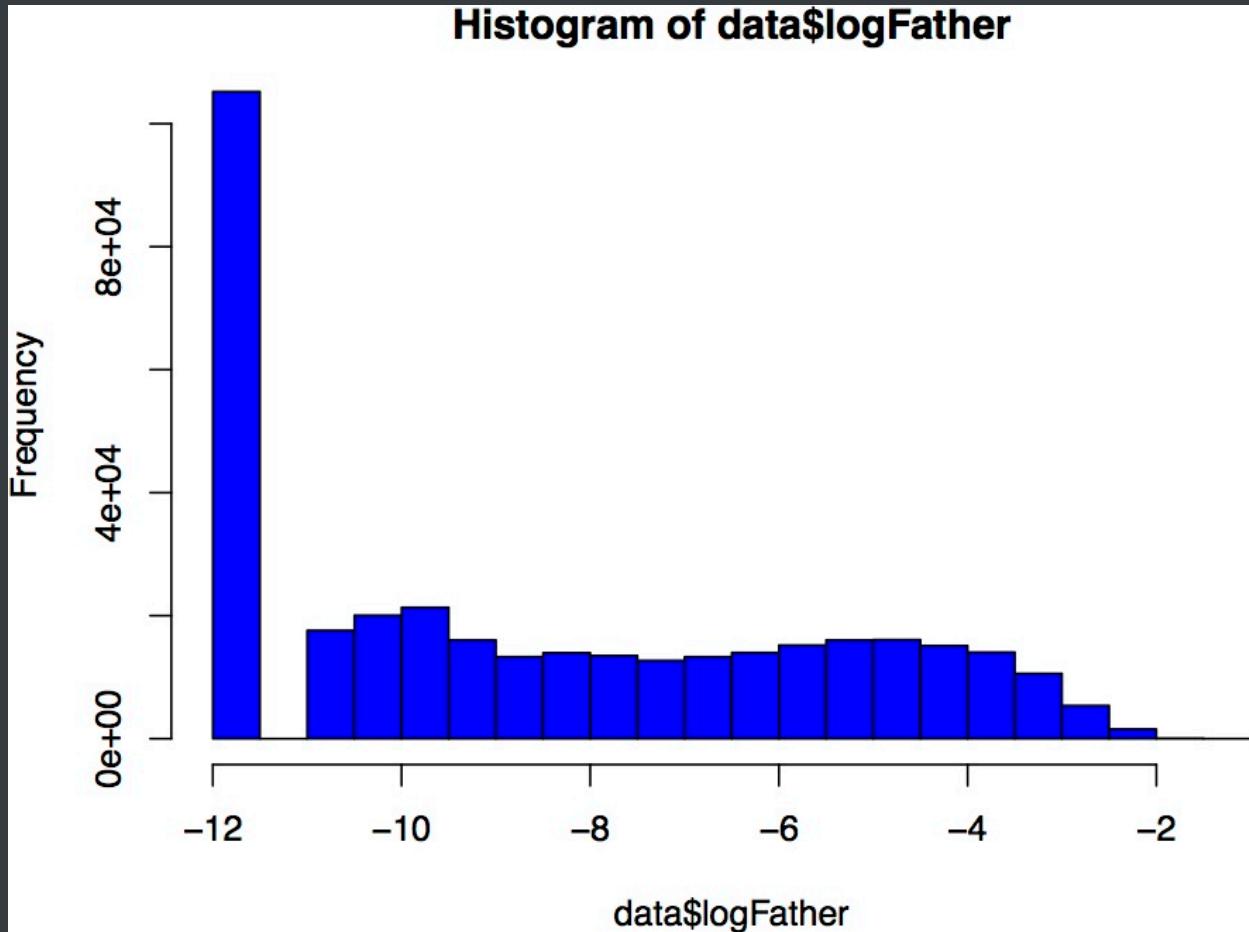
If we plot (in histograms and box plots) this raw data, we notice that the distribution is not normal at all. Very small motion history values are over represented and bigger motion history are much more rare with a very long tail. Consequently, it is difficult to compare two distributions from two different subject. To normalize the distribution to compare the motion distribution of several subjects and compute synchrony scores on it, we made the natural logarithm.

Unfortunately, several values equal to 0 and the log function can not be applied. They generate a -Inf value. If these values are set to NA, we lose the information of no movement at all. If we give an arbitrary value to this data by shifting the distribution by a very small value (eg. the half first value after 0), they are over represented) like in figure 2.



If we check all participants, we can see that the father and the mother motion history distribution are very similar. However, the therapist, which is always in a small window of the video with a different zoom, has a very different distribution. We have less signal on it. In some videos the patient is in this window, it explains, it intermediates position. Consequently, we can say that the signal extracted on the therapist is not directly comparable with the signal from the other subjects. A lot very small movements are not detected by the software and leads to 0 values. These values are over-represented.

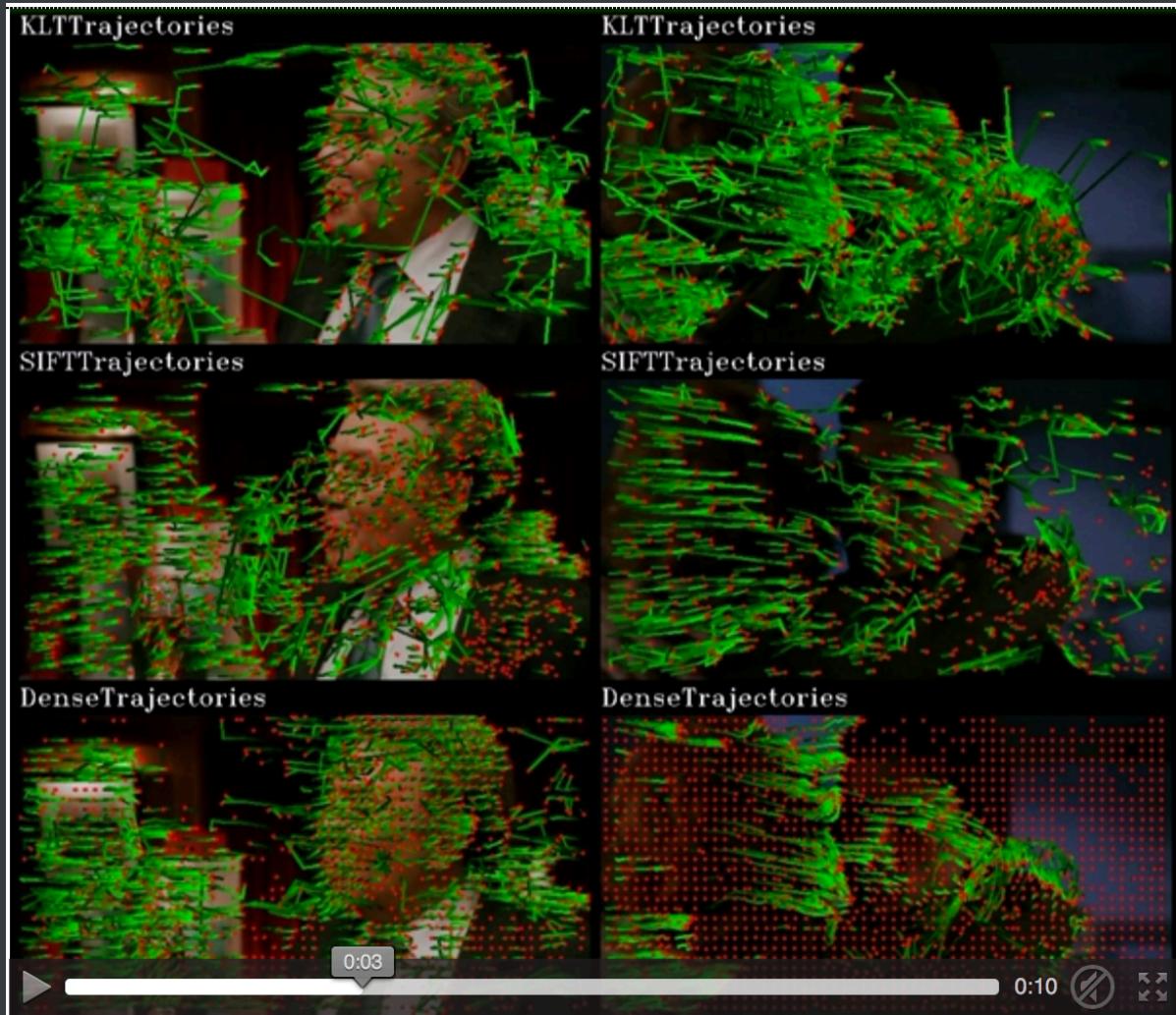




- It could be possible to extract background to improve the quality of the motion history. See for instance backgroundExtractor.py with openCV BGS library.

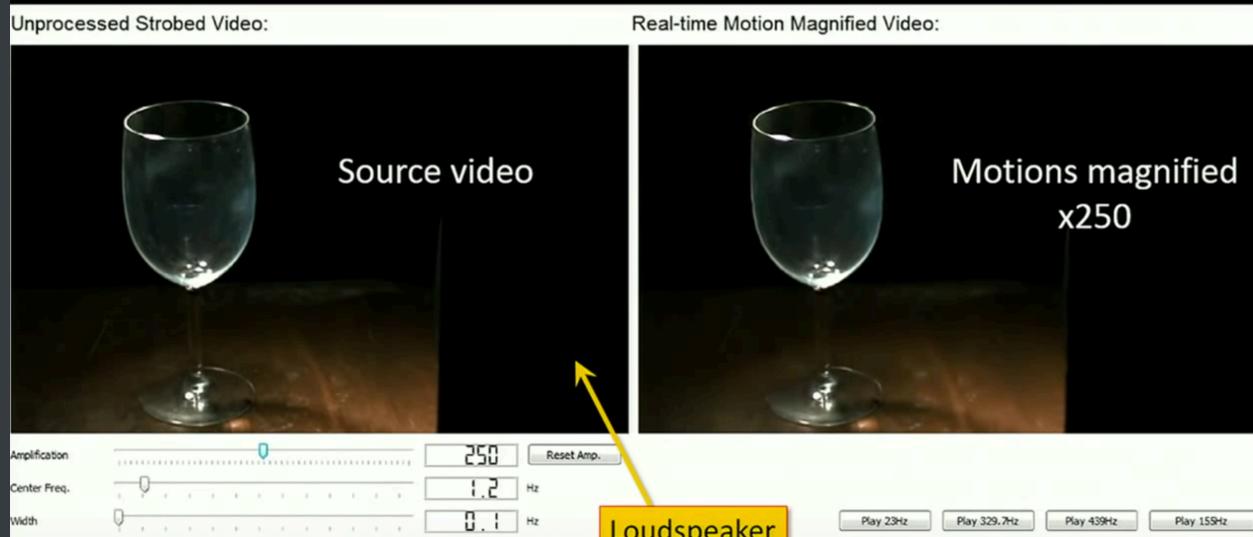
Instead of extracting only the motion history, we could extract more relevant signals.

- It could be interesting to evaluate the trajectories of the people. If they two people are going toward the same direction or not can change everything in understanding a human interaction even if the motion history is the same. For that, the optic flow modules of open CV or the Dense Trajectories Video Description could help.



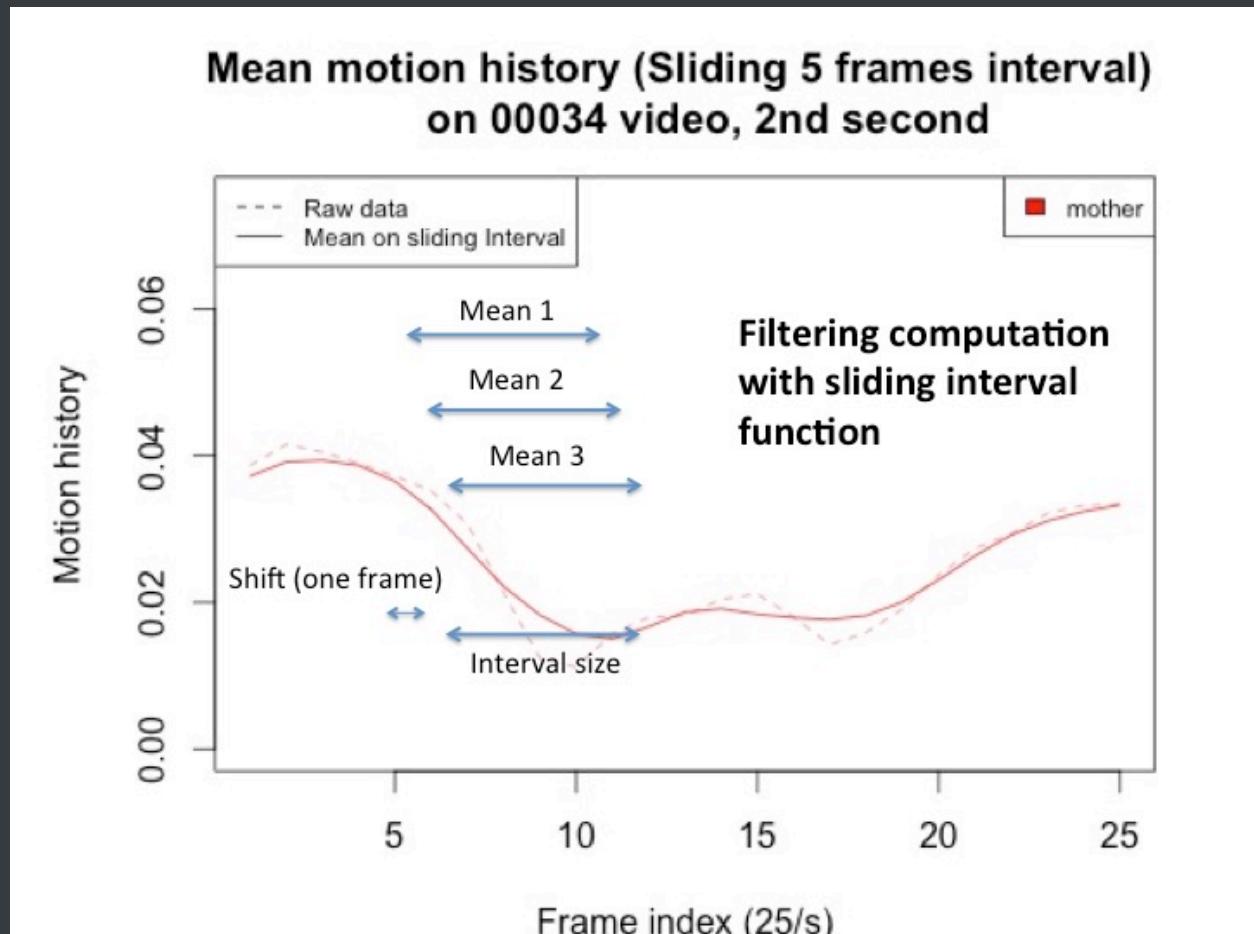
Tiny changes can't be always obvious. Some softwares can amplify a lot very tiny changes in specific frequencies that can't be seen normally with naked eyes. This enable for instance the extraction of the heart pulsation of the subject, which could be the reflect of the stress without any invasive physical contact.

Michael Rubinstein: See invisible motion, hear silent sounds. Cool? Creepy? We can't decide

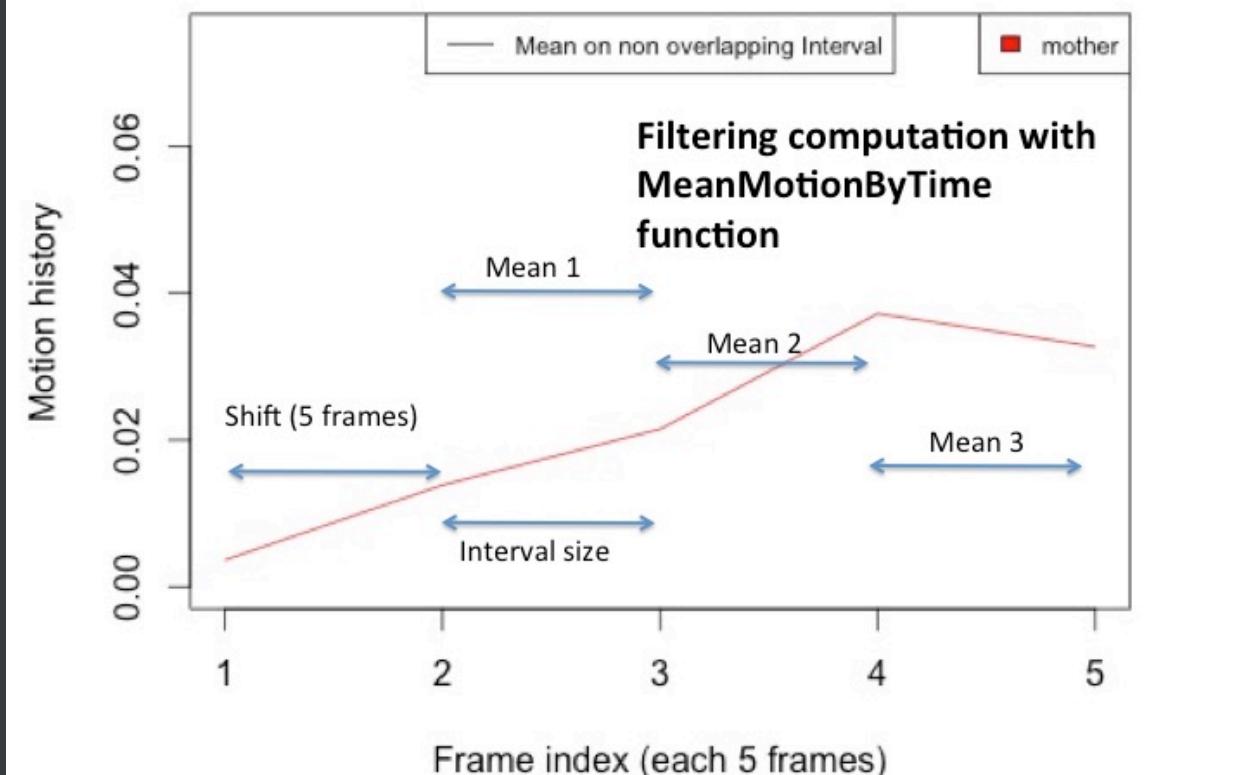


4. Motion history filtered with slidingInterval function

I filtered motion history raw data with a R function on Rmarkdown script analysis. It generates CSV file of the form "F1044C.VOB.slidedata.csv". See files [here](#). I developed two functions, one to filter motion history without changing the frequency of the data by sliding the reading frame to compute the mean.



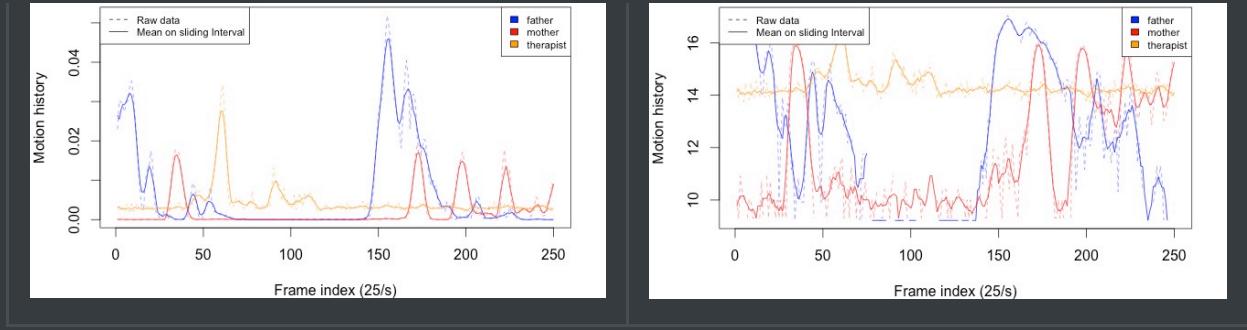
## Mean motion history (non overlapping 5 frames intervals) 00034 video, 2nd second



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
- idem with log TODO
- **NA** : corresponding to absent subject

Sliding interval (filtered raw Data)	Sliding interval (filtered log Data)
Mean motion history, (Sliding 5 frames interval raw data) on F1044C1 video, first 10 seconds	Mean motion history, (Sliding 5 frames interval log data) on F1044C1 video, first 10 seconds

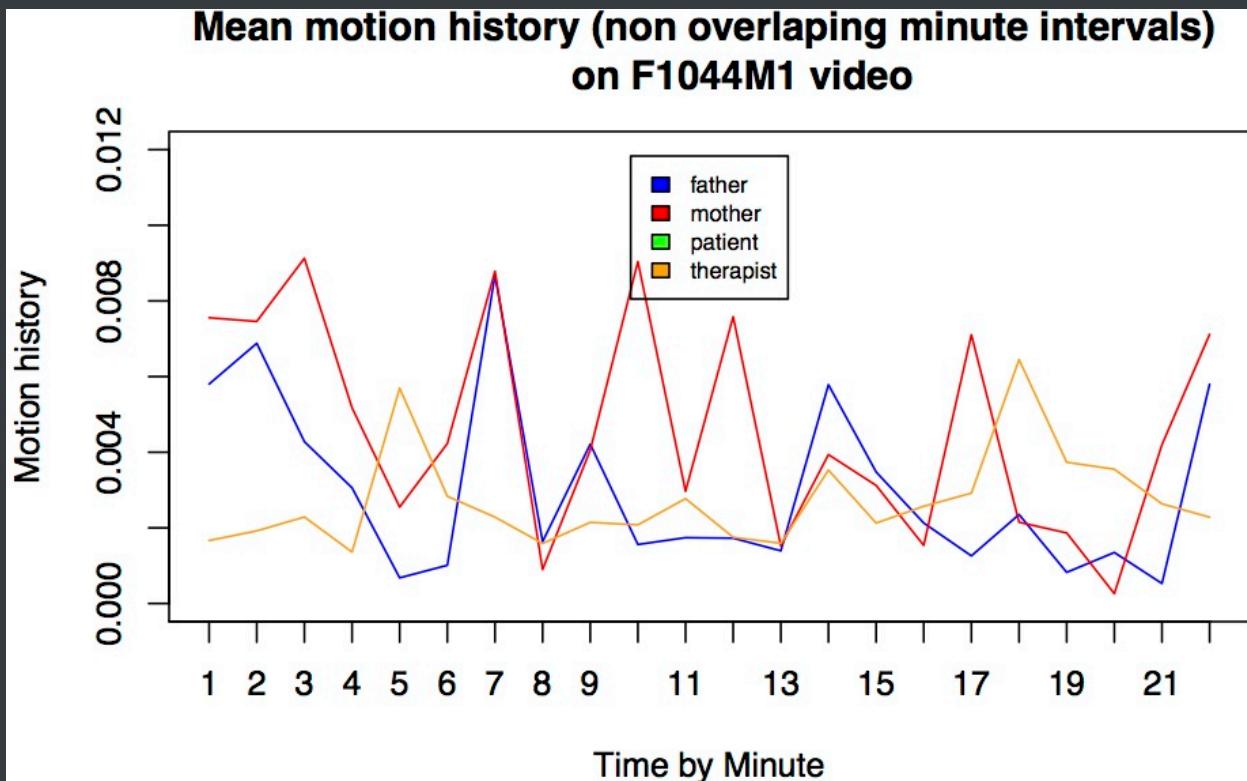


It is possible to change the size of interval to change the ration signal/noise.

We can see that there is a shift of the therapist motion history even in the filtered raw data. It is possible that the noise on this small window makes decrease the signal/noise ratio.

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.



The representation of the mean motion by minute enable to detect high change of motion history that could be the reflect of movement of all persons in the scene.

#### 4. Computing synchrony score with this filtered motion data

To compute synchrony SyncPy used the amplitude (difference between min and max). The baseline levels (min) are not relevant to compute synchrony.

In this configuration, the patient is embeded in the window, the mother and father are in front of the main camera. The therapist isn't filmed.

We downloaded and installed [SyncPy](#).

We adapted the script [Call\\_S\\_Estimator.py](#) that must be installed in Syncpy installation in examples folder. We used this script because we had continuous data and tried to compute synchrony between more than two subjects. We created a folder *SynchronyCSV* in this folder.

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.

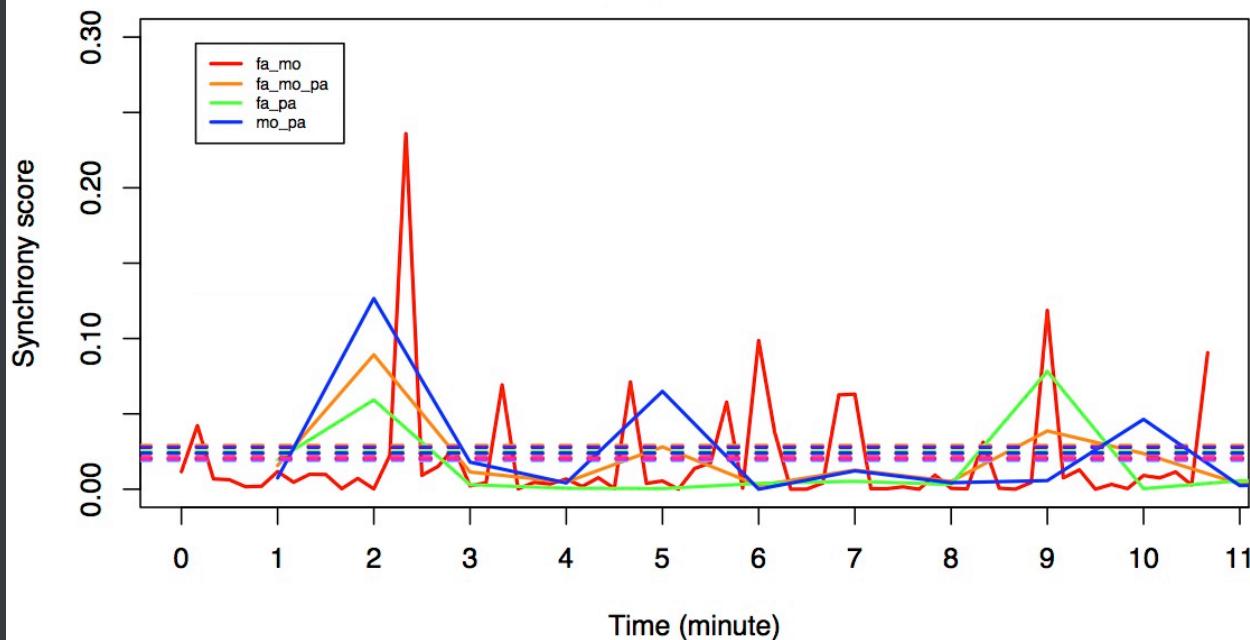
To compute the synchrony, we use the S\_Estimator algorithm that is drawn from information theory. It makes a matrix of correlation on the standardised motion history time series (mean =0 and SD = 1). Consequently, only the shape of the signal matters. Afterwards, it computes the eigenvalues of this matrix. Finally, after the normalization of the eigenvalues, it can compute the synchrony from a sort of entropy (Carneli et al, 2005). The index ranges from 0 (no synchrony) to 1 (full synchrony). If all signals are equals, the synchrony scores equals 1.

### **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 csv files are automatically detected from this folder and S\_estimator is used on them.
- the number of frames used in an interval to compute synchrony, Nota bene, there are 25 frames by second. (eg. `numberOfFramesByInterval = 25*10`). We chose this parameter since it seems to be clinically relevant but other parameters could be tested.

This process is relatively long and can take several minutes, especially with short intervals. It returns synchrony CSV files that can be found [here](#).

### Synchrony scores for each dyad and for the whole group in F1044D1 video



We expected that the signal to fluctuate between 0 and 1. However, we can see that the signal fluctuate only between 0 and 0.25 with a lot of synchrony scores very low around 0. The synchrony scores are better with the subjects in front of the main camera i.e. the mother and the father in this configuration.

We saw that the distribution of motion history was much more informative when we compared the log of it instead of the raw data. This enable to detect the aberrant distribution of the therapist and the periods with no motion at all which were over-represented.

When we did the log scores before computing even if it was interesting in order to compare motion history distributions, however, that didn't change the relevance of the synchrony signal.

When we tried to check what could be the relevance of this signal, since the psychotherapy wasn't very operationalized, it was difficult to match the two. We contacted the INCANT team however, we couldn't meet them quickly.

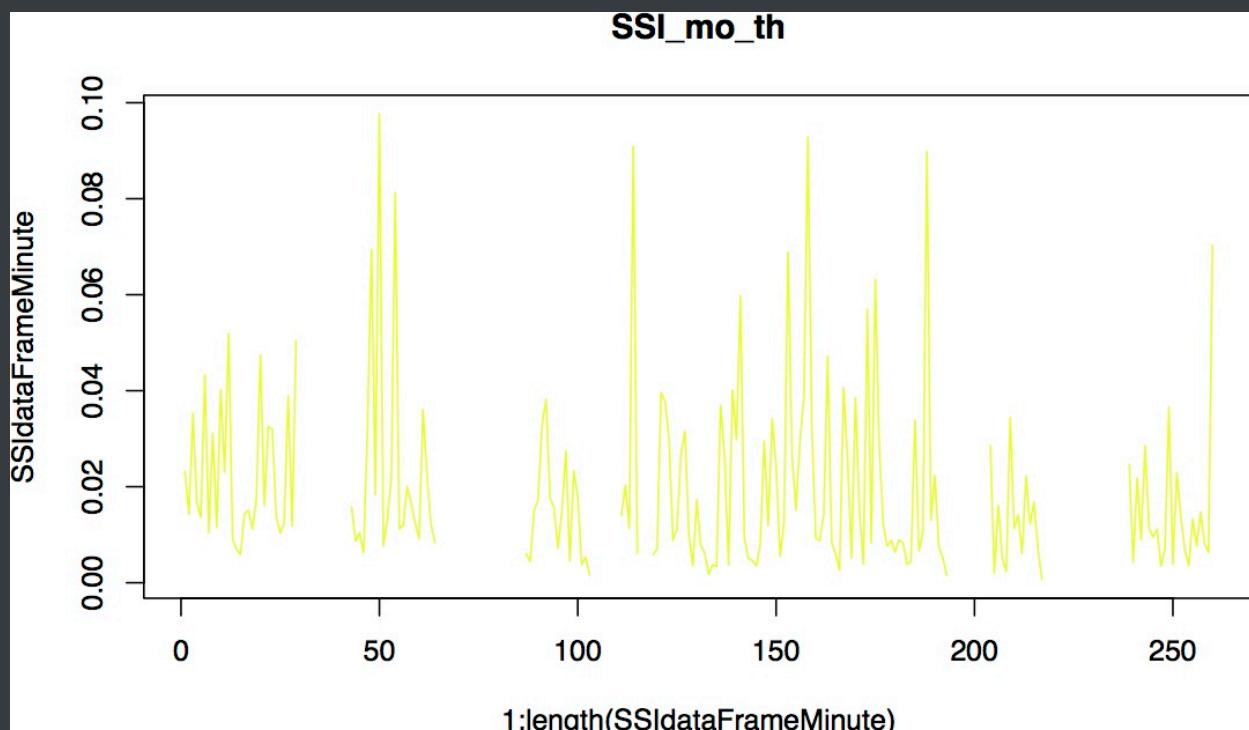
Headers :

- **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 all, 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).

We can evaluate the evolution of the synchrony for a dyad or tryad through the evolution of the psychotherapy. Here it is represented hrough the whole therapy. Sometimes, there wasn't any data because the configuration didn't enable that. We can notice that is not obvious to evaluate and compare the synchrony of a dyad if it is alone or with other participants since the interaction could be different.



We saw that the evolution of cannabis use of F1044 is not straightforward. It is quite high at the beginning, decreases and then increases. Consequently, we couldn't evaluate of the dynamic of the synchrony could be a predictor of the psychotherapy and the consumption of cannabis since this ast dynamic wasn't clear.

We decided to select 4 subjects with a clear decrease of cannabis consumption, we expected that this better dynamic could give clearer results.

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

All of them with the same therapist Mrs Bastard. However, we had problems with the conversion of the files from VOB to AVI. The number of videos from this families were much lower, to evaluate a dynamic could even more challenging. Since, we got a better database with less technical problems on it and a experimental setting, we decided to switch to it to further developp the concept.

### Advantages of this database

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.

### Limits of this database

#### Experimental set-up

It is however not very defined in the articles. Consequently, it was difficult to contrast synchrony between two well defined periods. A much more operationalized psychotherapy like motivational interviewing could be useful. This kind of psychotherapy can be rated with the degree of adherence of the psychotherapist to the principles of the therapy like the MITI. Consequently, we could compare two quantitative variables that will reflect the quality of the psychotherapy at a more precise moment instead of a more general outcome.

#### Technical limits

It would be better to have a more experimental set up to improve this problems, 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.

It seems that it would be very important to have several devices (camera, kinect, mikes, actimetr). This leads to a problem that wasn't raised in this set up since there was only one camera. In case, of several devices, it is very important that all of them are synchronized.

To begin to work in a multiperson psychotherapy was theorically possible and the team developed scripts for that on SyncPy (S\_Estimator) however, it is much more challenging that evaluate the synchrony of the dyad since the configuration needs to be comparable, the subject mustn't overlap and must be filmed in the same conditions (zoom).

### Conclusion

We couldn't answer the questions with the INCANT study we asked in the pre-registration, because of the technicals and quality problems of the database and experimental setting. We can't say if the motion history synchronization between different participants are correlated with the outcomes of the familial psychotherapy (decrease consumption of cannabis).

Since the number of videos can be very low, we couldn't neither say if the synchrony could be the reflect of the adherence.

However, this database can precisely enlighten some points that need to be addressed to consider this evaluation.

## MONRADO study

### Aim of the study

The aim of the study was to evaluate the response to a conflict in a dyad of a parent and a child. There were two phases. After the experimenter left the room, the dyad spoke about casual topics. After around 5 minutes, the experimenter came back in the room and asked the dyad to speak during around 10 minutes about a conflictual conflict that was selected by the child before the video during an interview alone with him or her.

Consequently, from these videos, we can extract easily two phases that we could contrast : without then with conflict.

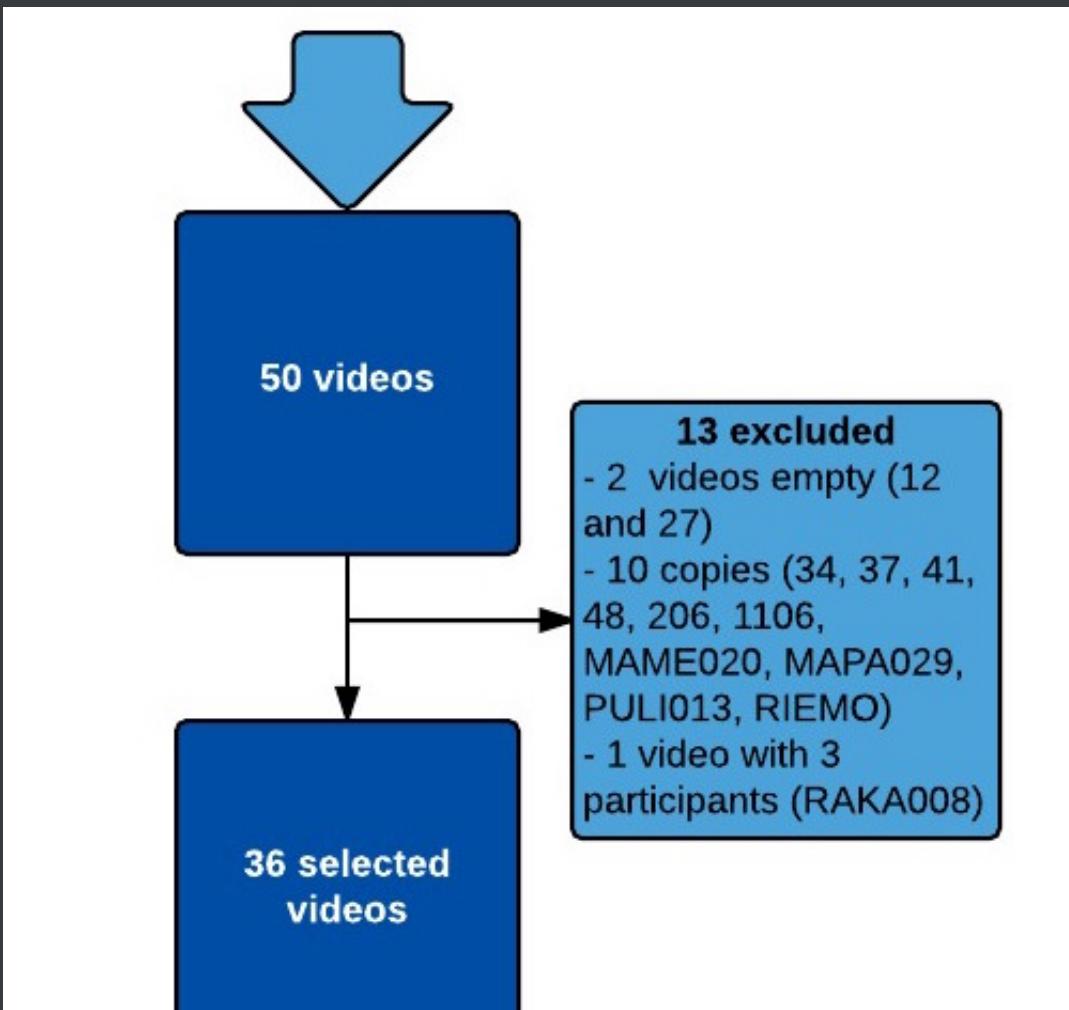
### Configuration

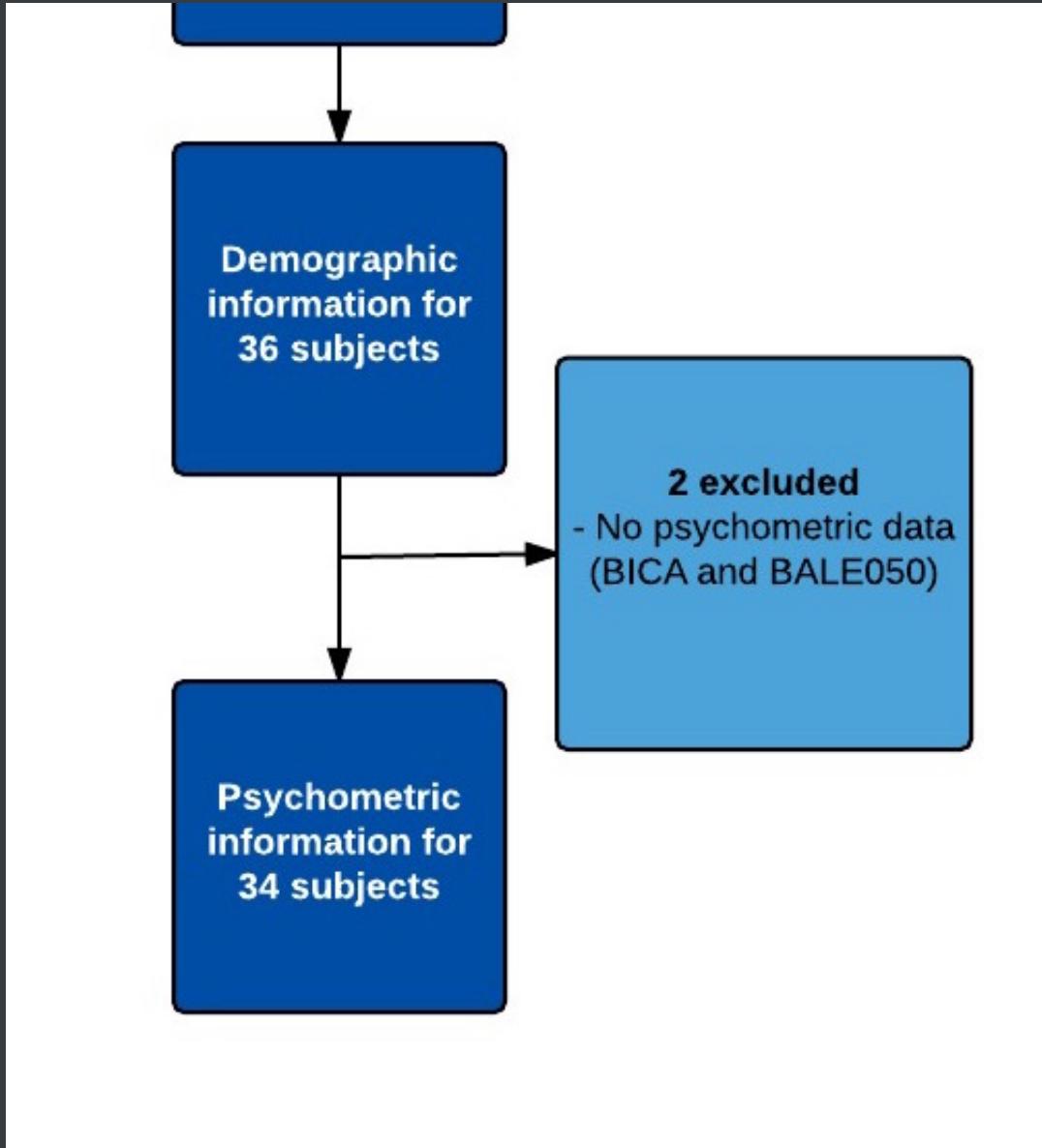
The participants are two in front of the camera. We can see that the quality of the video is much better. There is not often overlap between the subjects. There isn't embedded subjects with a different zoom level. There is not any date embedded. All the subjects that takes part in the interaction are filmed.

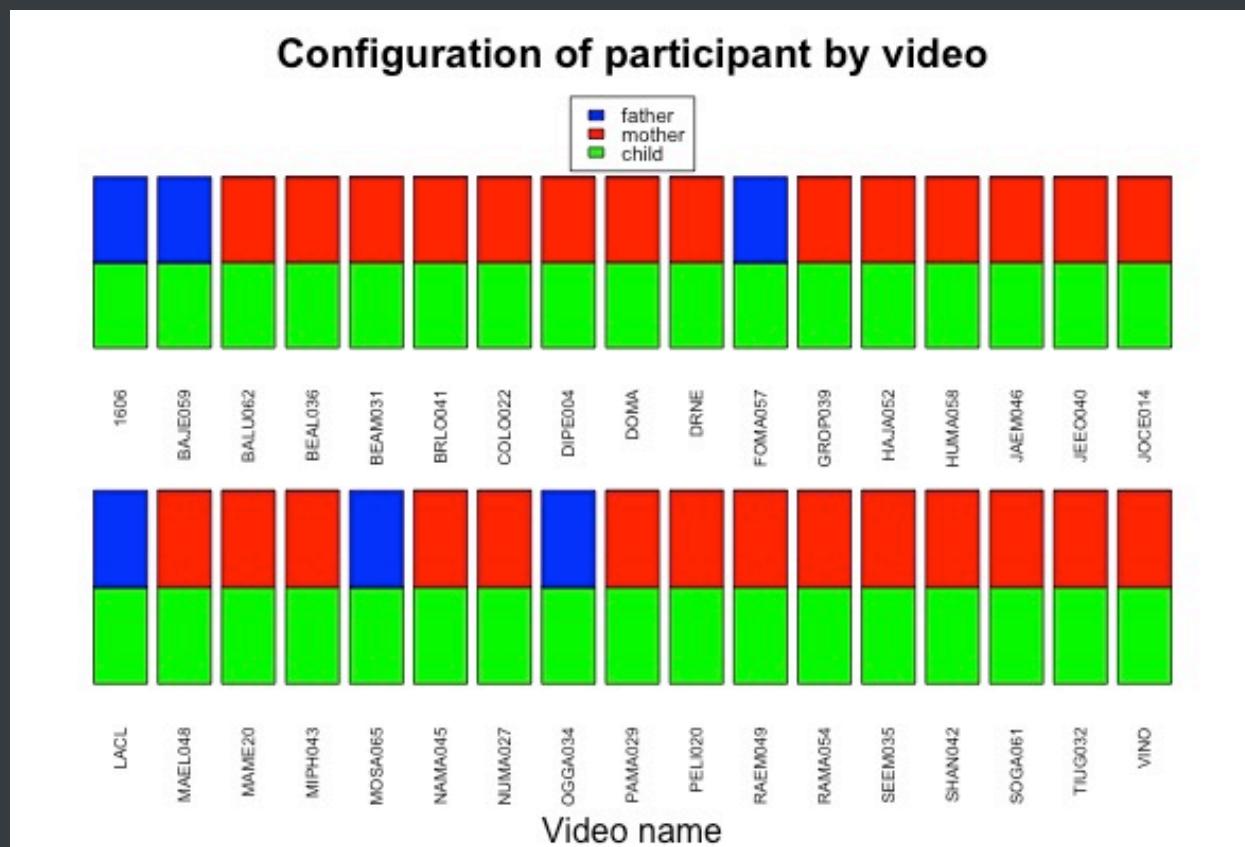


### Selection of relevant videos

In order to match the psychometric data which was more important and the videos, we decided to clean the database to delete the copies and keep only the data with psychometric informations. We decided to delete a video with a different configuration with the two parents.

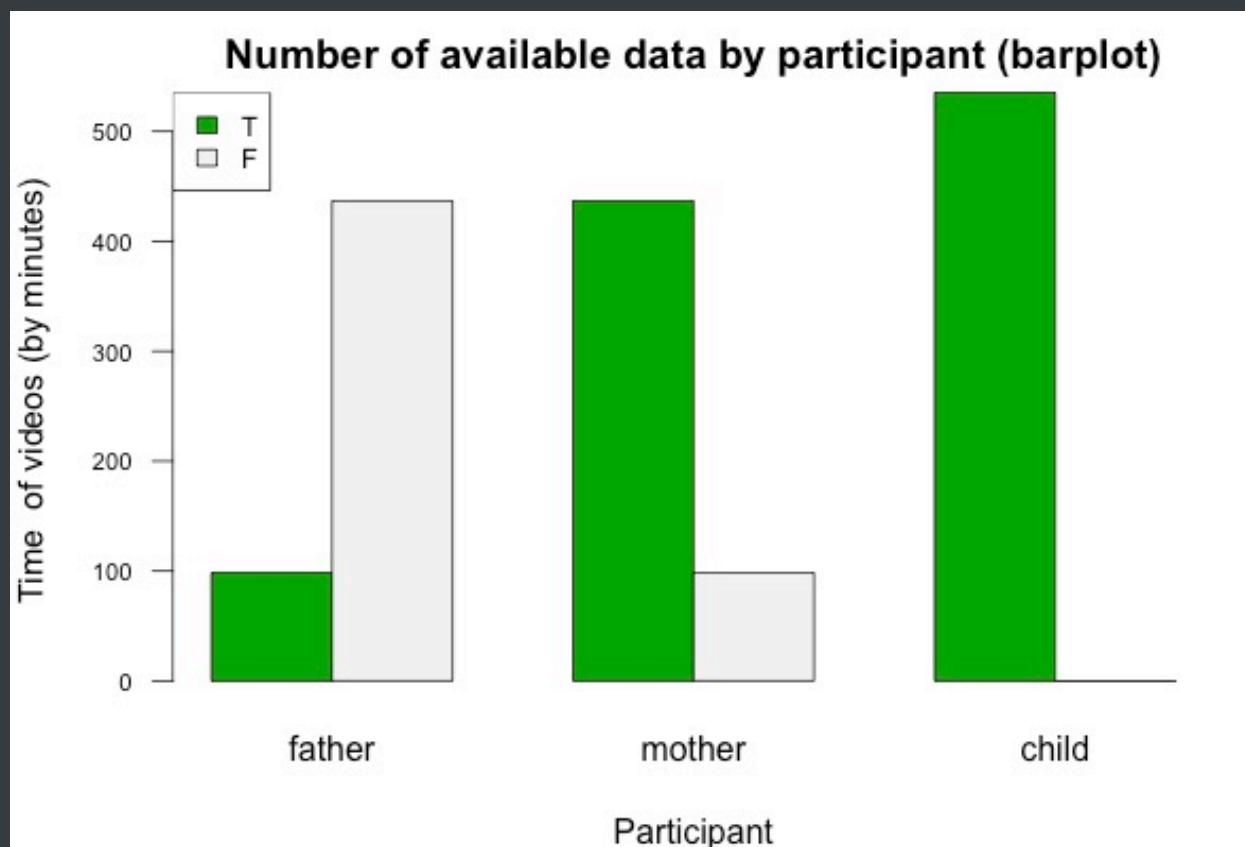






Presence of participants,

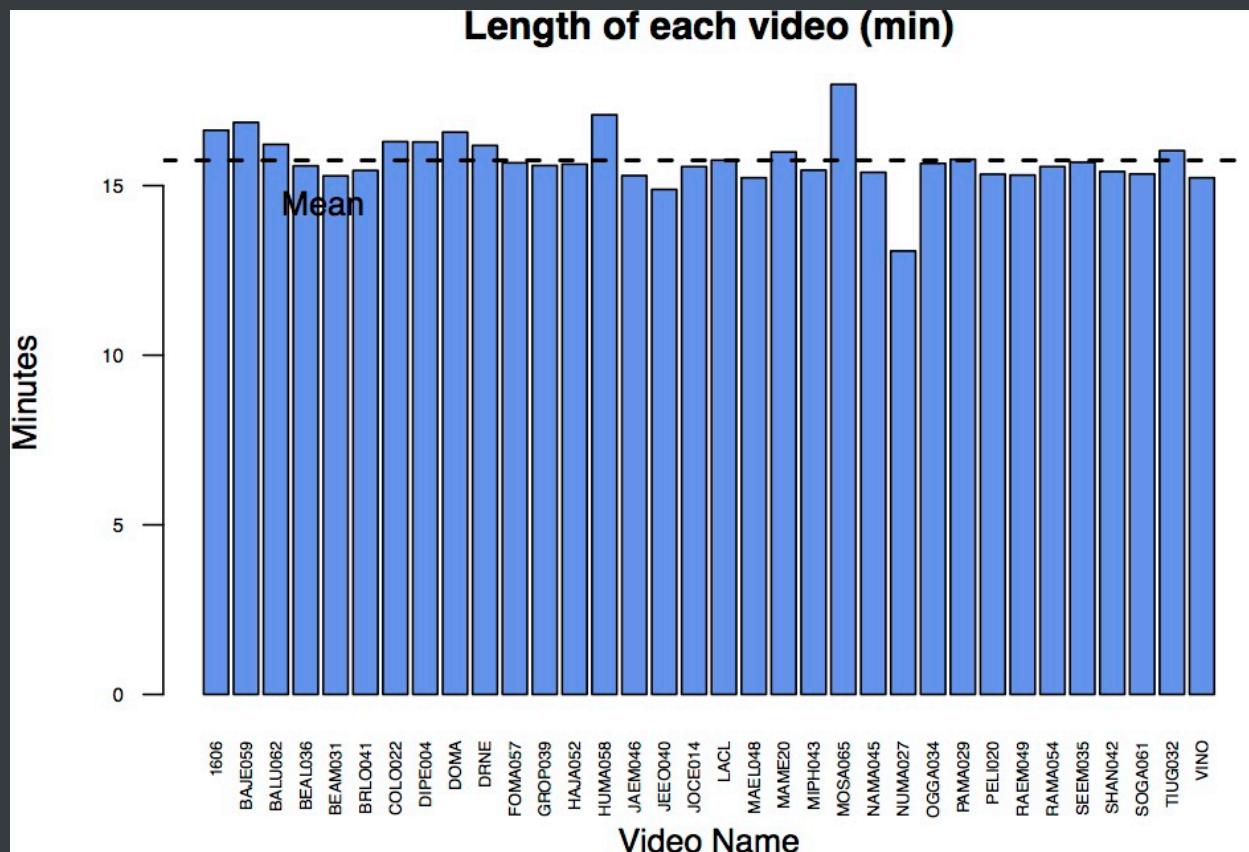
The configuration in the different videos are very similar.



In these videos all the participants are filmed. The child is always filmed. More often, the mother is involved but not the father.

TODO nombre de minutes

Length of the videos

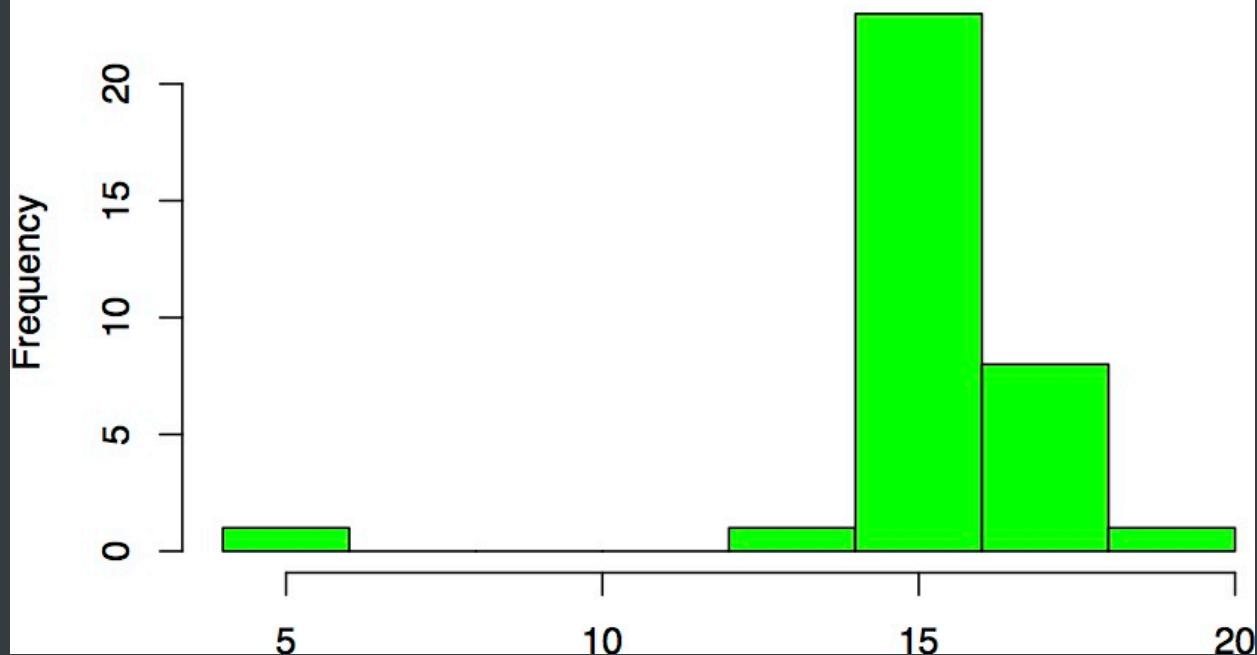


We can see that the length of the videos are very similar. In NUMA027 video, the instructions were not very well understood by the subjects. The mother left the room before the end to see the experimenter.

Psychometric data

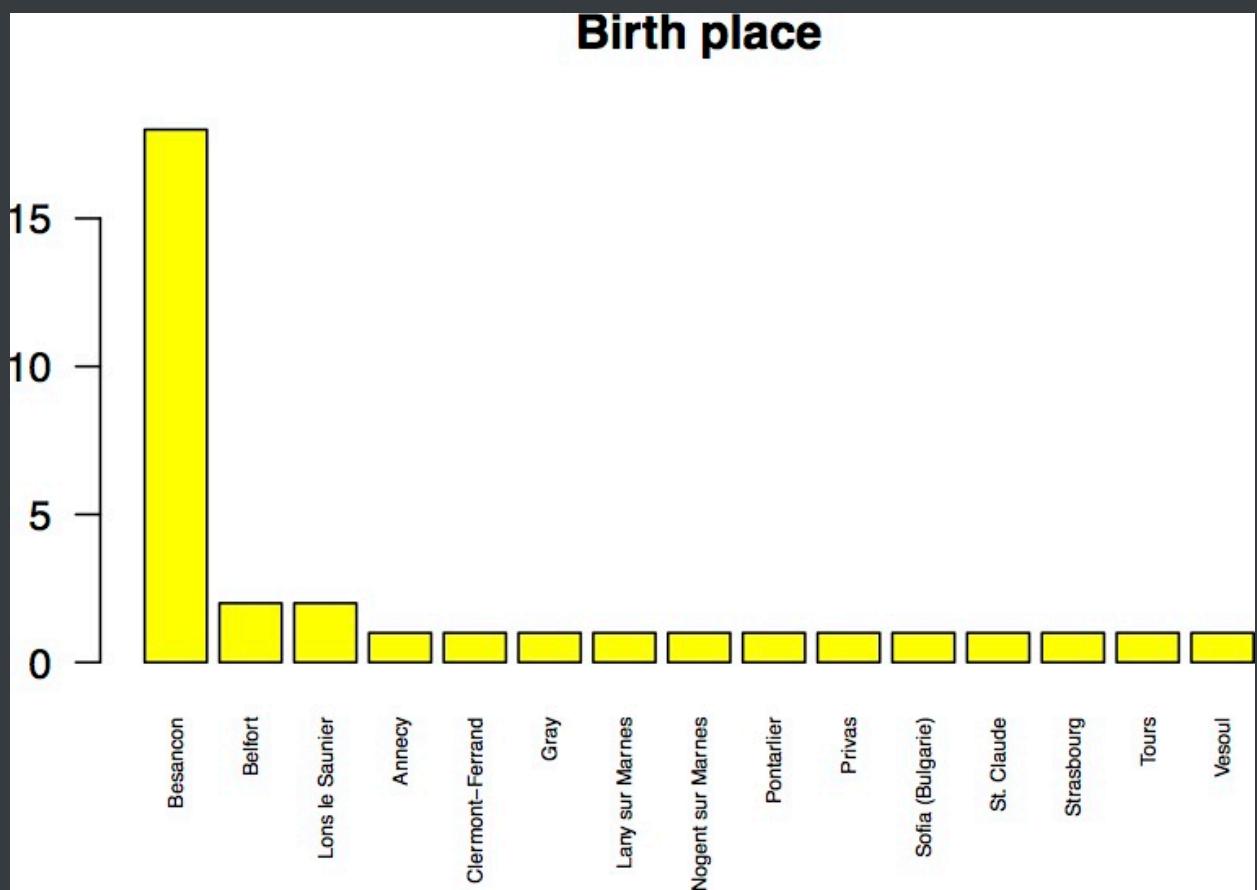
- It consist of one excel file with:
  - Identification number
  - Identification of the video
  - Demographic information (Age, Birthday, Birth place)
  - Attachment style
  - TAS Score
  - STAIYA Score
  - BDI score

## Histogram of the age of the teenagers



The mean age of the participants was 15.4 years (SD : 2.0).

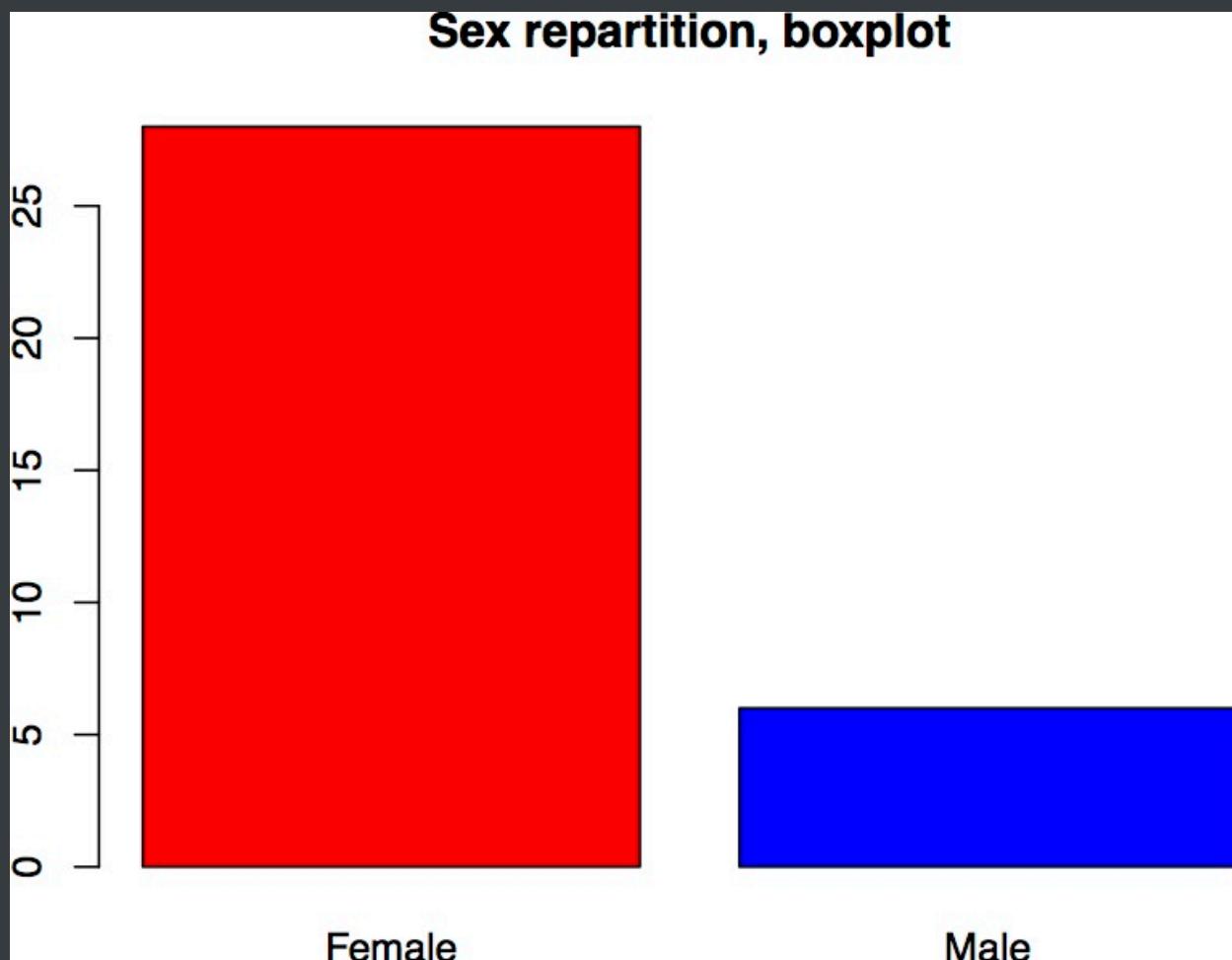
## Birth place



Most of the children were born in Besancon.

TODO Frequency

Sex of the child



82 % of the teenagers involved were female.

TODO FrequenCY

Attachment style

Attachment theory is a psychological model that tries to describe the dynamics of short and long-term relationships between humans.

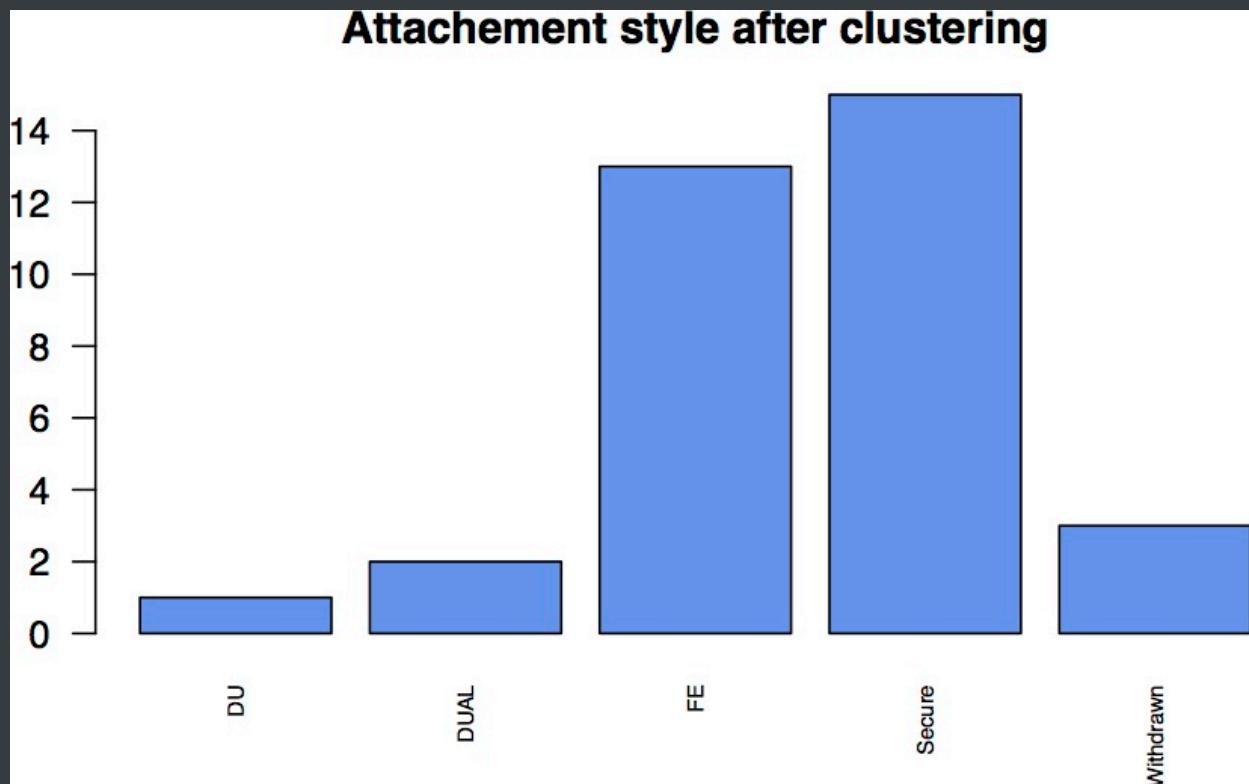
The Attachment Style Interview for Adolescents (ASI-AD) is a user-friendly, transparent and reliable measure. It offers the adolescents the opportunity to talk about their recent life events, their close relationships (with parents and friends), their support networks and how they feel about getting close to others. (Jacobs et al, 2012). Caring about attachment in young people in residential care: The use of the Attachment Style Interview. ('Community Care Inform' electronic information source. Spring 2012)

Catherine Jacobs, Yael Ilan-Clarke and Professor Antonia Bifulco, Lifespan Research Group, Centre for Abuse and Trauma Studies) <http://www.attachmentstyleinterview.com/>

Attachment styles were Secure, Enmeshed, Fearful, Withdrawn and Dual when two different styles)

After a contact with the MONRADO team, we clustered the attachments styles in 3 groups:

- Secure
- Fearful-Enmeshed (FE)
- Withdrawn
- DUAL were excluded



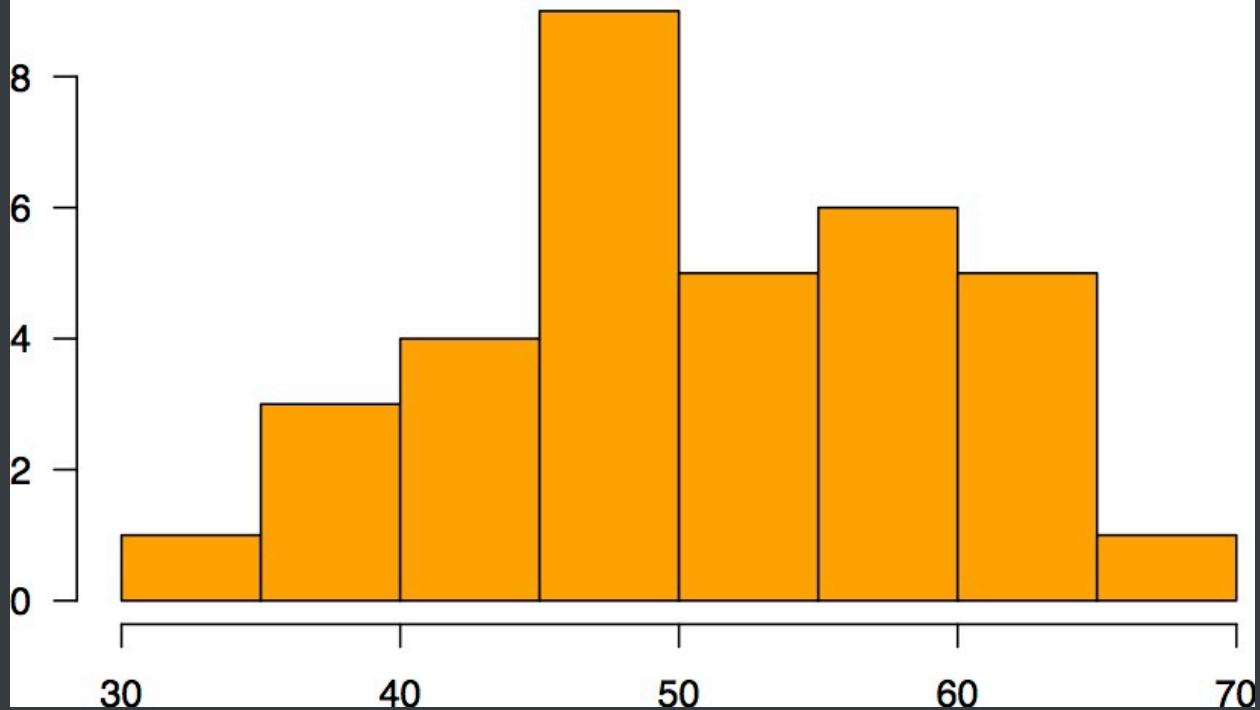
Secure (44%) and Fearful-Enmeshed (FE) (38%) clusters are over-represented.

#### TAS Score

##### The Alexithymie

The TAS (Toronto Alexithymie Score) that was developed by Taylor, Graeme J.; Ryan, David; Bagby, R. Michael (1986). "Toward the development of a new self-report alexithymia scale". Psychotherapy and Psychosomatics, 44, 191-199. It was translated by M.P. Marchand et G. Loas (1994). It evaluates the level of Alexythymie. If this level is high, the ability of the adolescent to understand and express his own emotions is low.

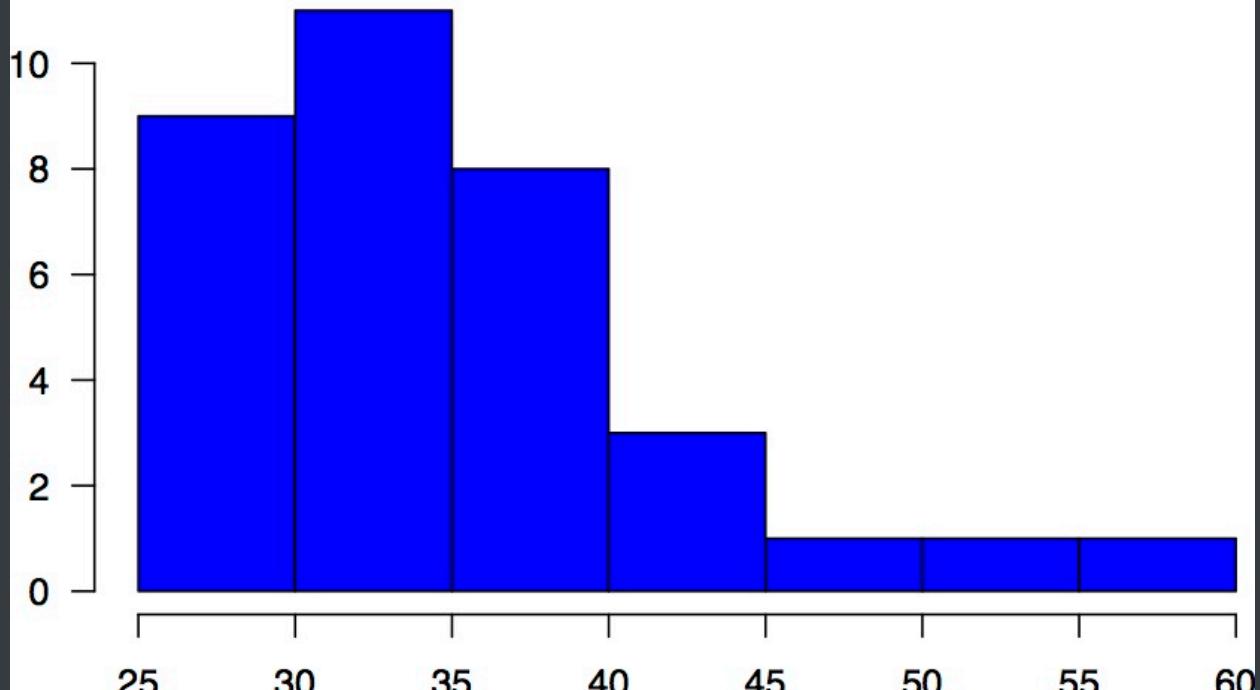
## Histogram of the TAS scores



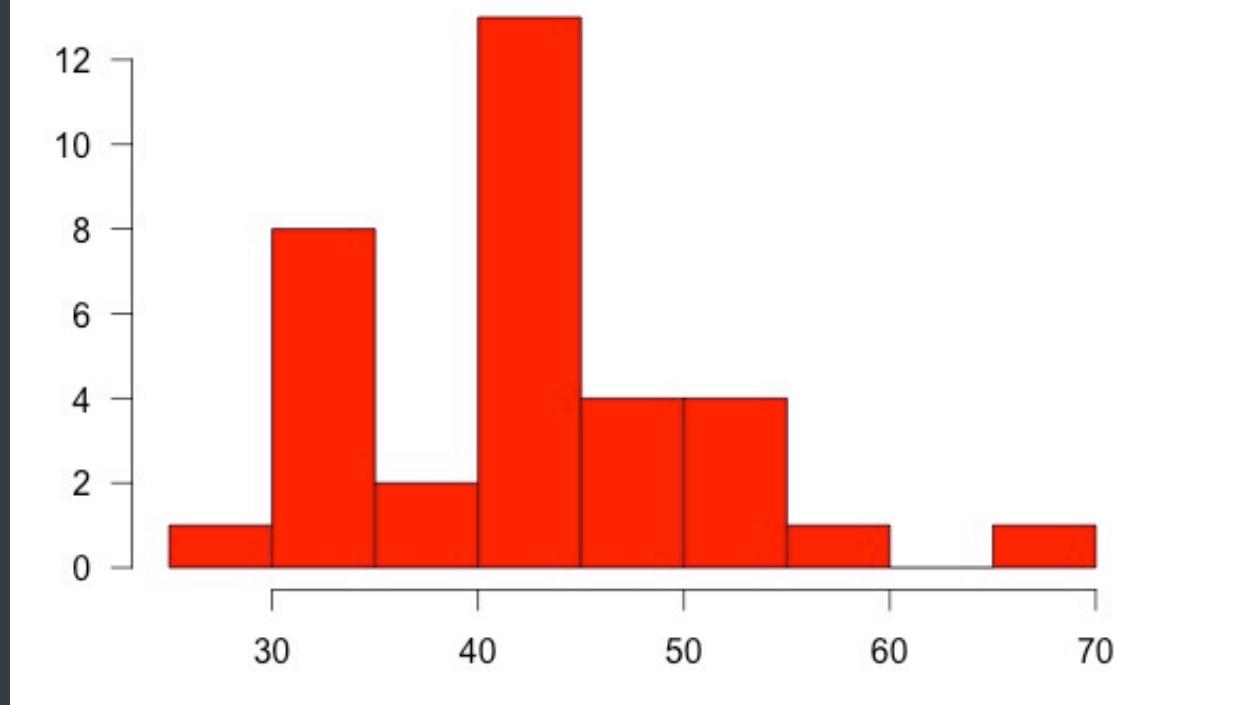
STAIYA Score

The STAIYA or State-Trait Anxiety Inventory of Spielberger is an auto-evaluation of Anxiety. It is one of the most used Anxiety auto- questionnaire. It dissociate the anxiety at a precise moment (state) and the habitual anxiety of the subject (trait). When this quetionnaire was developped, it was evaluated on more than 5000 subjects. The score ranges from 20 to 80.

## Histogram of the STAIYA scores



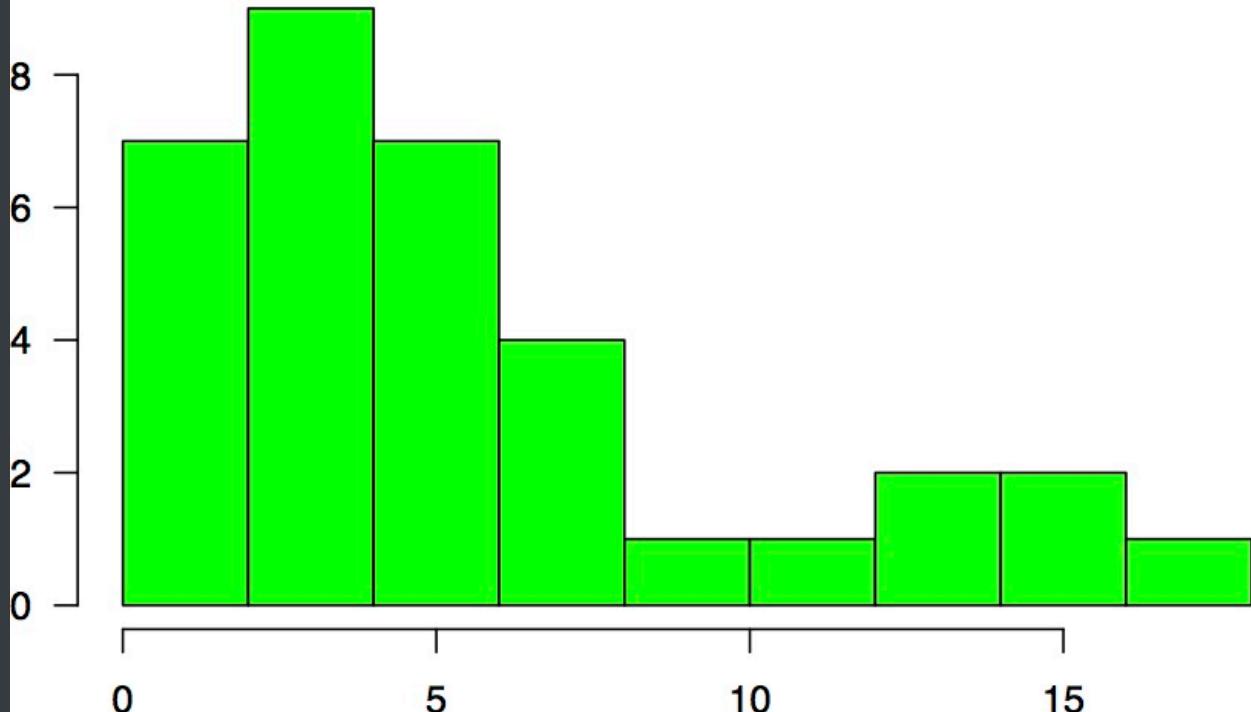
## Histogram of the STAIY-B scores



### BDI Scores

The Beck Depression Inventory is a depression score ranging from 0 to 63. It evaluates the severity of a depression when it was clinically diagnosed. From 0-13, the depressive state is minimal, between 14 and 19, the depressive state is mild (Osman, 2008). This measure supports that the teenagers recruited in this study are not severely depressed people but not depressed people.

## Histogram of the BDI scores



### Description of the database

- 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.

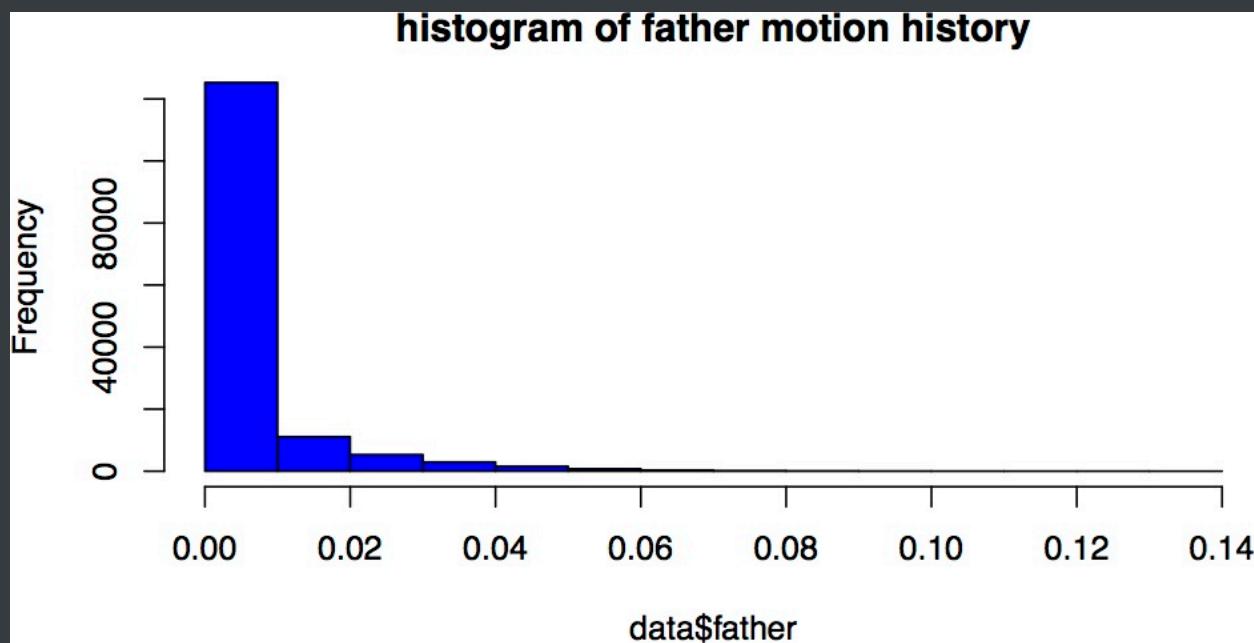
We used the same procedure that for the INCANT study, we masked the non relevant part of a frame for each subject. Instead of selecting only the upper part of the bify, we were able to select the half of the image with more information about the subjects.



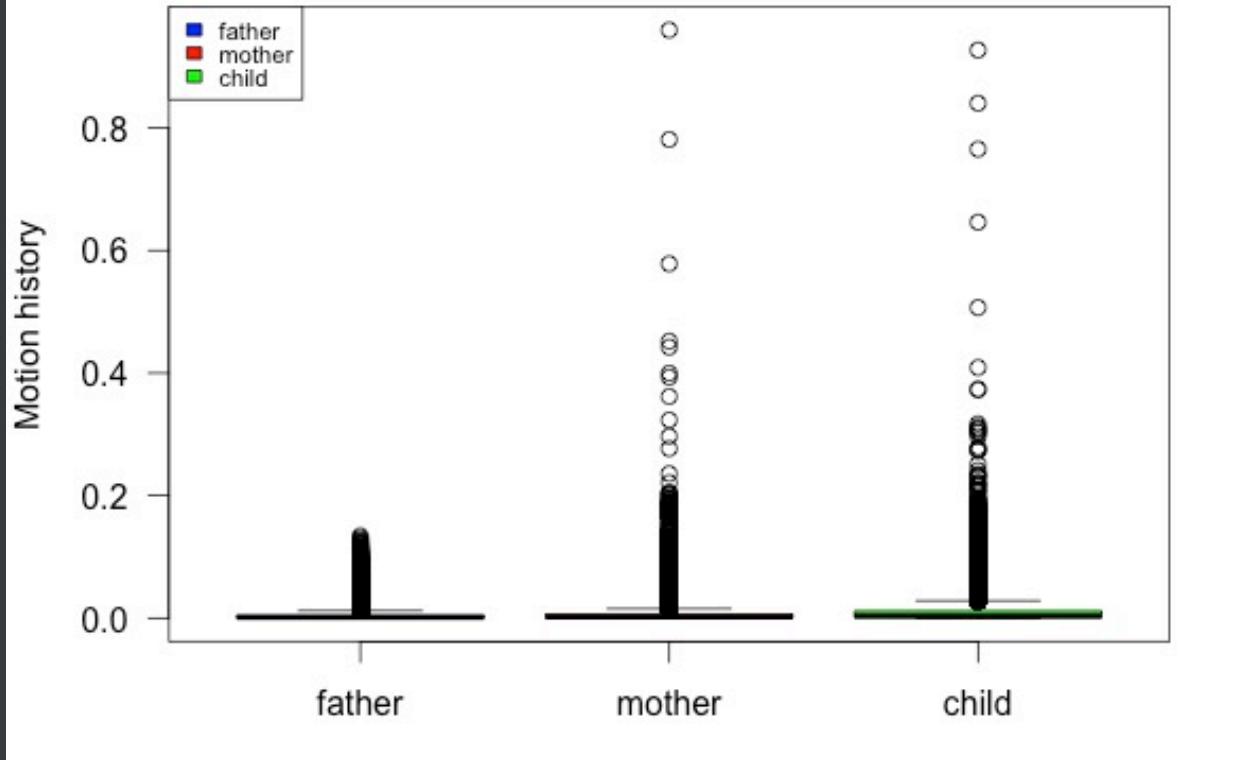


Extraction of motion history

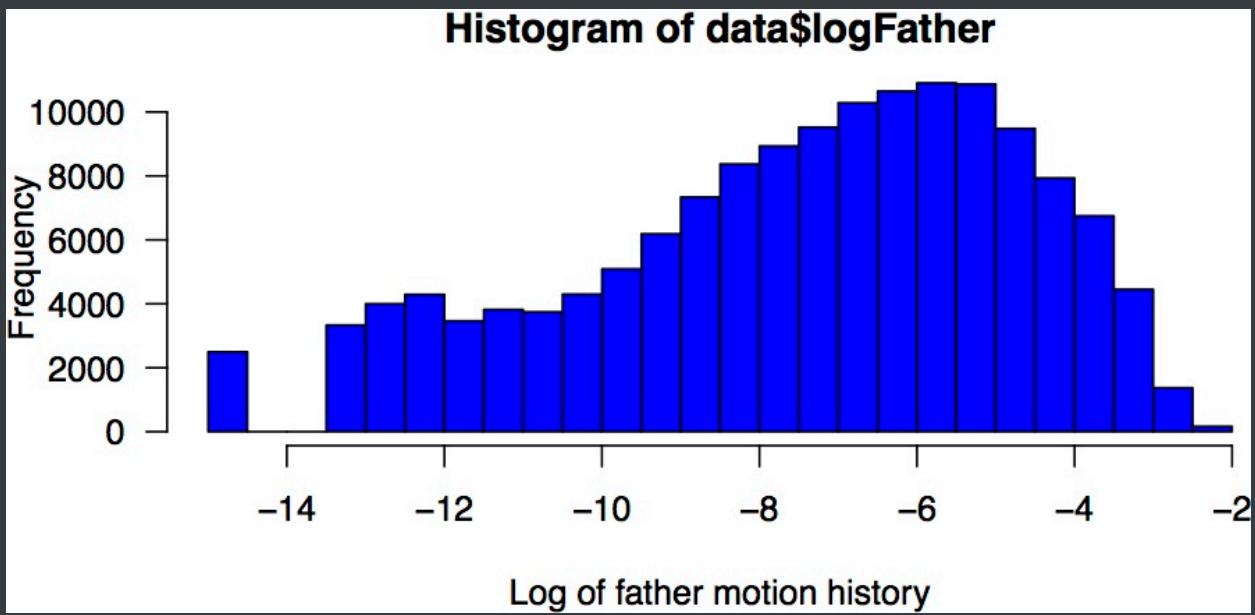
TODmother



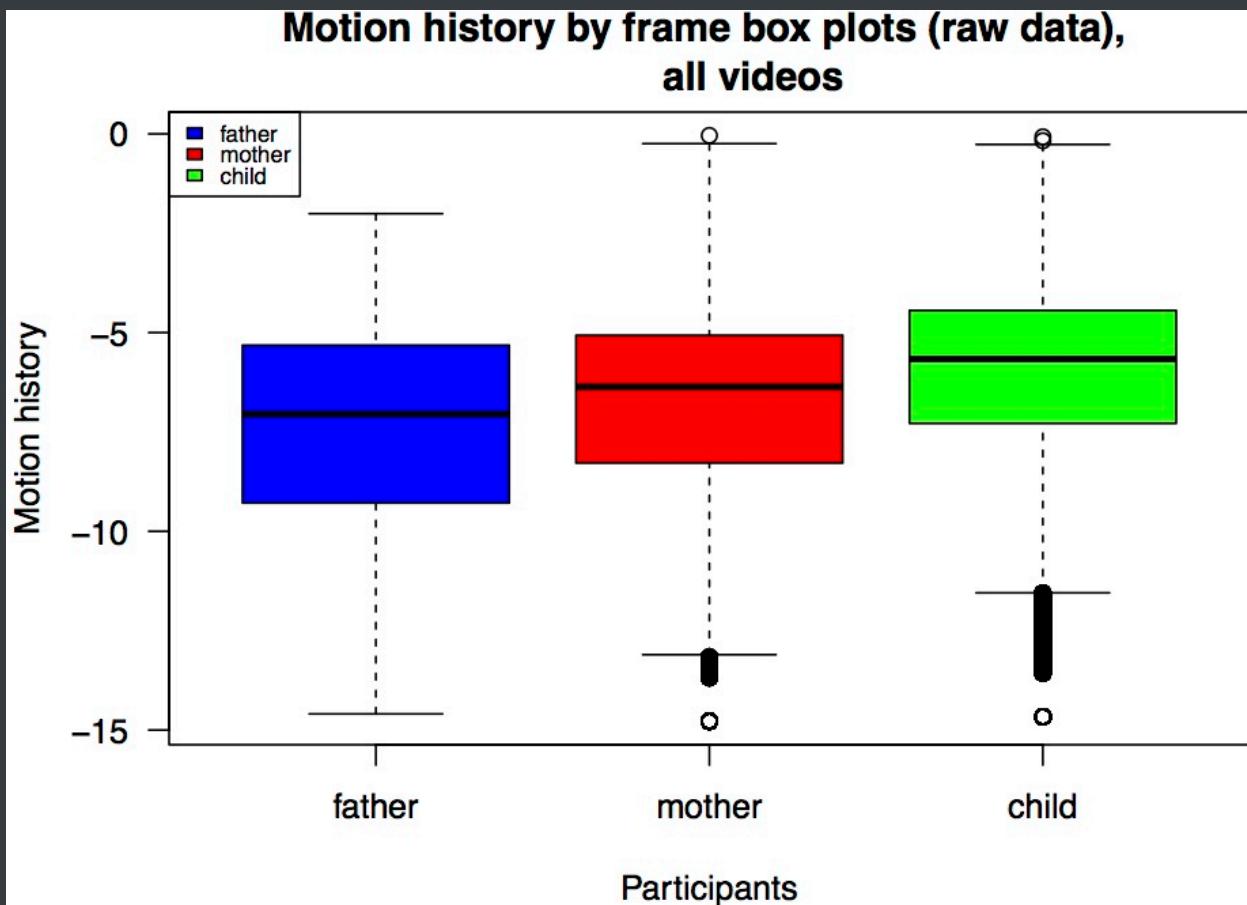
## Motion history by frame box plots (raw data), all videos



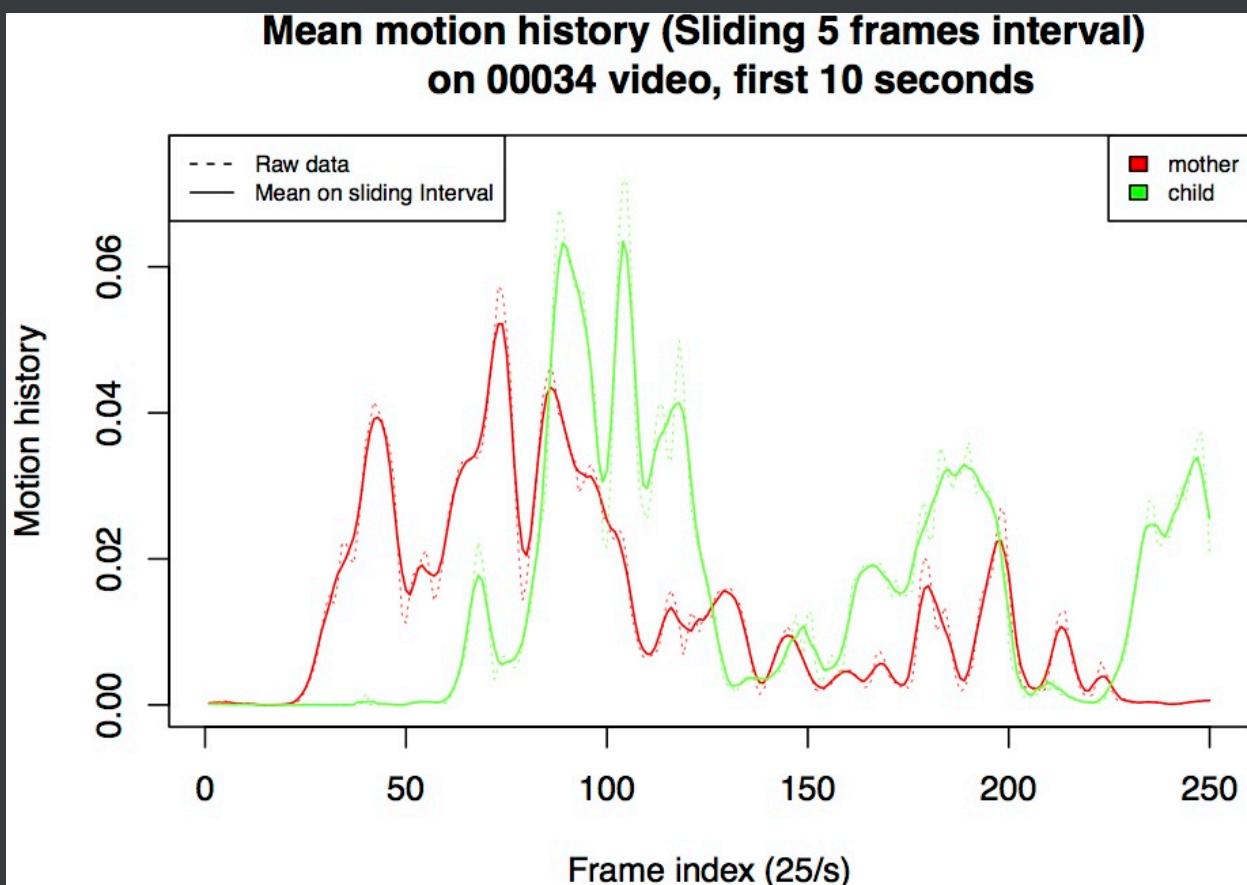
We can see that the distribution of motion history is very similar with the motion history extracted from the INCANT study. The small movements are very important and the big movements are very rare with a long tail.



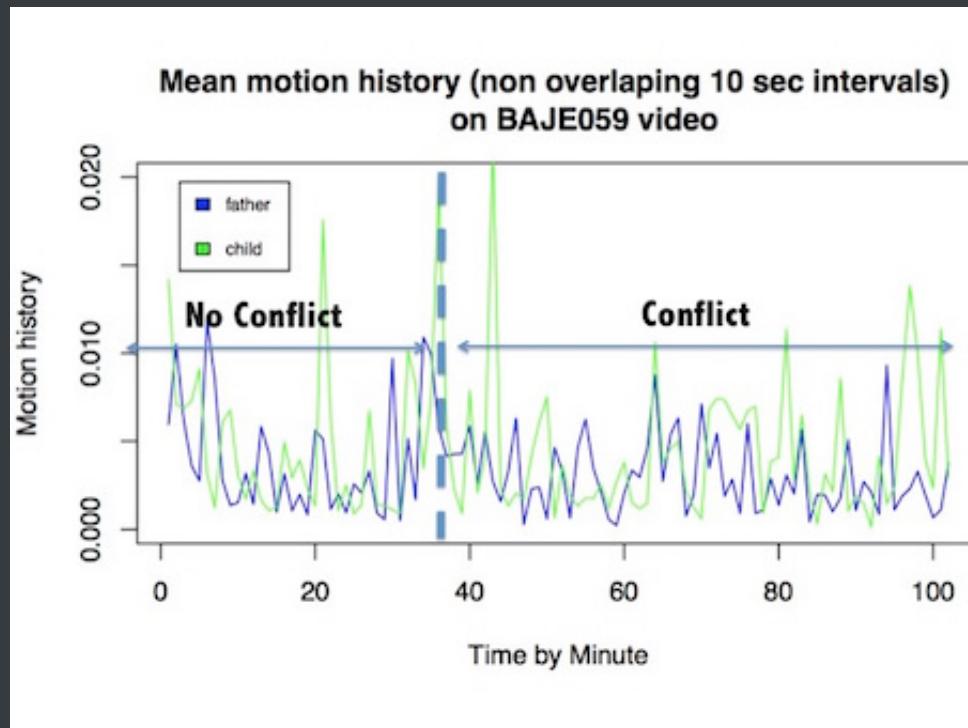
When we compute the log of this data, we can notice that the distribution is much more normal than for the INCANT study. The first bin representing the null motion histories (no motion) is much smaller and not overrepresented. This confirms that the quality of the motion history signal extracted from this video is much better than the previous one extracted from INCANT study.



When we compare these distributions, we can see that these distribution are much more similar than the distributions in the INCANT study in which the therapist motion history distribution was very different from the other distributions.



Like in the INCANT study, we filtered the motion histories.



TODO : cut zone in the middle

We could very easily detect the two periods of the video. The first one without any conflict and the second with a conflict. These periods were annotated manually.

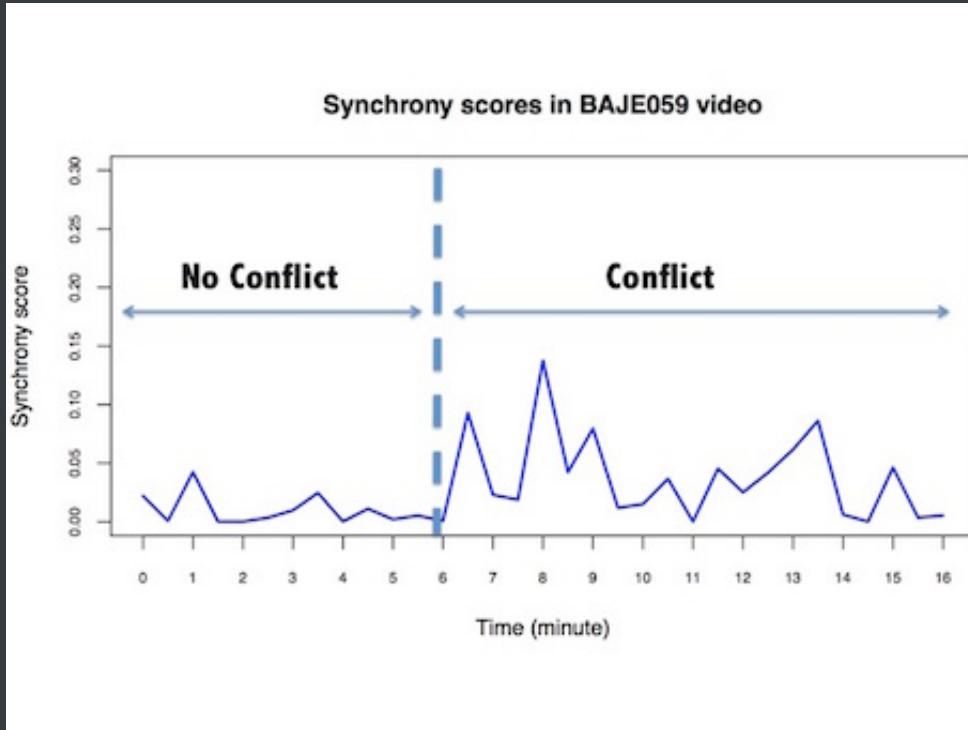
The cutFrames data frame was done manually by looking manually all videos and defining:

- when the experimenter leaves the room and the interaction begins,

Between is the non conflictual discussion

- when the experimenter comes back to ask participants to have a conflictual discussion
  - when the experimenter leaves and the conflictual discussion begins
- Between is the conflictual discussion
- when the experimenter comes back to shut down the camera

However, we can notice that the conflict wasn't always very authentic and important. It would be interesting to ask outside observers, blinded with the objective of the study to annotate the level of the conflict of the participants. That would be done with crowdsourcing and will be much more easier than the annotation of speech that we did for the INCANT study.



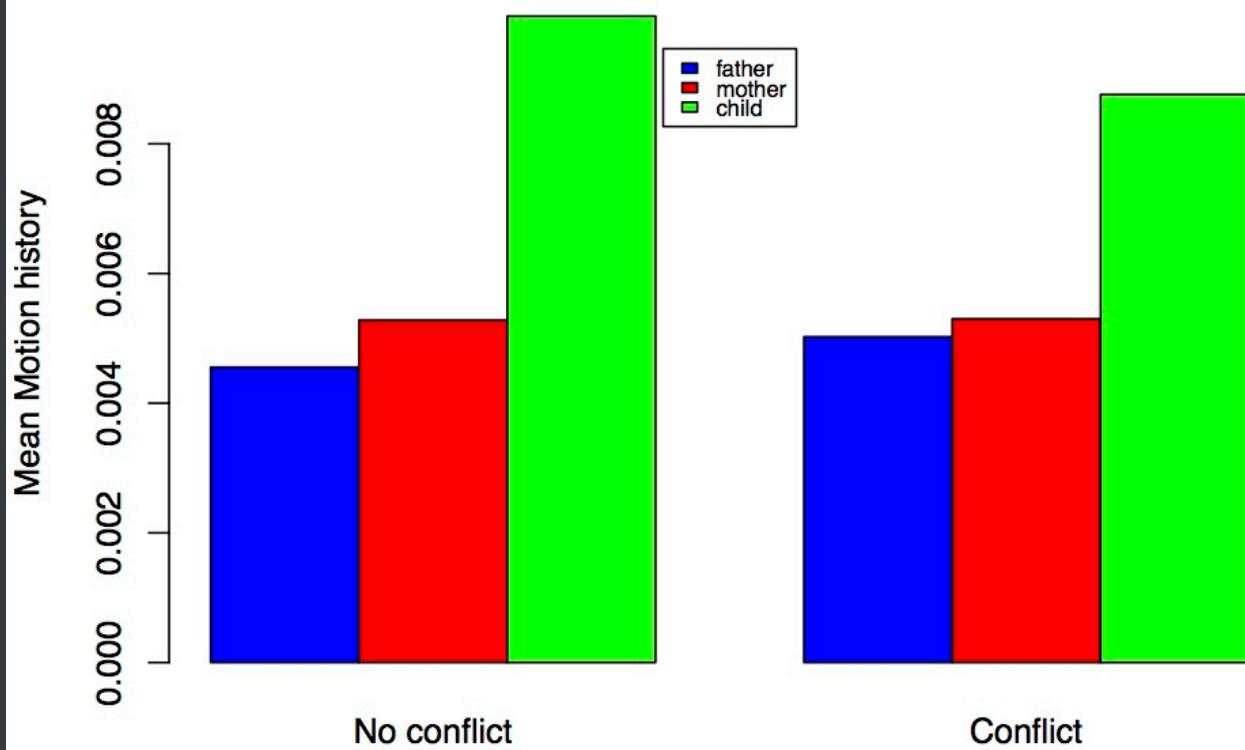
TODO : cut zone in the middle

However, we can see that the synchrony score are not so much different that from the INCANT study. Although, it was expected to fluctuate between 0 and 1, the maximum of it is around 0.2-0.3 depending on the videos. When we did the log scores before computing however, that didn't change the relevance of the signal.

Thanks to that, we were able to compare the mean motion history between the two conditions.

t-test

### **Mean Motion history for each participant without and with conflict**



### **Comparaison between SSI without Conflict and with Conflict**



As expected and like previous authors showed before, it seemed that the synchrony is lower in conflictual situations.

TODO t-test

AUC?

coorelation entre données psychometriques et SSI

The next steps will be to evaluate the synchrony scores in different conditions (different psychometric scores, especially attachment evaluation) and make proper statistical evaluations. We consider too, presenting the process to psychotherapy training team to get raw data of better quality. We will meet a representant of the Motivational Interviewing Network of Trainers (MINT).

### Technical problems

None

There was only one video which no movement of the camera. There wasn't any date emebeded in the videos. There was very few overlap of the subjects.

However, the subjects weren't always looking at the camera that would be a problem to automatically detect their emotions, their smiles for instance.

### Results

#### Discussions

It could be relevant to annotate the level of conflict by blinded annotators since the conflict is not always completely obvious nor strong.

## Discussion

#### Experimental discussion

We saw with the INCANT study that there was a limit with contrasting different conditions. In order to make this possible, it would be usefull to use much more opperationnalized psychotherapy like (Assertiveness, Motivational interviewing, exposition), but it is not restricted to CBT. We could use it in (provocation vs no provocation used in brief psychotherapy, hypnosis induction versus no hypnosis induction, dream analysis )

#### Analysis discussion

We could make a shuffle of the data to check if the signal is more synchrone than expected by chance. We could try to replicate this data with a more simple algorithm : correlation.py which was specifically designed for a 2 persons synchrony study.

In the future, we could integrate much more kind of social signals and the cluster of it in a more general score could be more relevant. Some techniques in modelling used at the moment in neuroscience in fMRI could be very usefull for this purpose.

# Perspectives

We could extract other features :

## ▪ Psychometric

- **Annotation** : annotation of conflict and non conflict period with Elan software to get better time resolution especially with more complex annotations. In subjective annotations if they are not too much difficult, it would be usefull to use crowdsourcing annotation by asking naive subjects about the level of conflict or stress that can be detected on a video.

## ▪ Physiological

- **Cortisol** : Cortisol levels can be the reflect of stress. Some techniques are minimally invasive like saliva measures(Teruhisa et al., 1981).
- **Oxytocin** : Ocyctonin could be involved in attachment and trust (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005). This could improve the social competence of Asperger patients with some social impairments. We could imagine to evaluate the relationship and the synchrony of two subjects normal and with social impairments with and without oxytocin.
- **Cardiac Frequency** : direct or indirect on videos with systems like Lambda or webcam pulse detector
- **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 information

We could be much more precise in the measure of the motion by using a Motion capture devices. There are basically two systems:

- We could measure directly very cheaply motion by using actimetry devices that looks like watches on the wrist. Unfortunately, that could help for motion of hands but that couldn't give precise information about the motion of the head.
- One is minimally invasive with a Kinect, a special camera that project infra-red light and can detect it to measure depth of a scene. It has a much a higher frequency signal detect that can rebuild a 3D representation of a subject and extract its skeleton. We saw that the low frequency of the signal we got 25 Hz could be a limitation of the synchrony signals we could get.
- Optitrack that are some markers on the body of the subject that can help to create a 3D representation of the subject.
- An other strategy is to use special suits like in the Ethome project that could record precisely a lot of motions directly.
- **Facial / Upper body recognition** : We could use automatical detection of Facial / Upper body recognition to segment images automatically instead doing it automatically.
- **Emotions**: We could use a camera for each person to have a more precise vision of its face to detect emotions more easily.
- **Background subtraction** could be useful to improve the quality of the motion we get
- We could measure the **gaze direction** to check if the quantity of gaze interaction could be a

reflect of a good relationship like it is described widely by the psychiatrist and psychotherapist in social impairments (social phobia, autism, schizophrenia)

- **Trajectories** : We could measure optic flow and trajectories instead of motion since the
- **Speech extraction**
  - We could anticipate the problem of speech extraction by using mikes to compute the synchrony of this signal

## Bibliographie

<https://www.zotero.org/groups/448197/items>