

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

**Author: Thomas GARGOT**

## **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 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 already feasible to annotate manually videos. However, this task is tedious. 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) and a potential source of bias. In the future, an automatic feedback during the psychotherapy could also even help the therapists to reorganize a treatment.

In the future, an automatic feedback during the psychotherapy could also even help the therapists to reorganize a treatment.

Here we suggest a research framework to extract automatically social signals from psychotherapy videos. We focus on motor synchrony considered as a predictor of psychotherapy outcome in a first study (Ramseyer, 2011) and a relevant signal for the study of mother-child interactions. 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.

### **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 and to identify the challenges of this process.

**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 automatically 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 to categorize easily the patients in good or bad responders and to detect specific periods of the psychotherapy that could be of particular very good or very bad relationship 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 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.

## Results

This work led to the development of R scripts to manipulate motion history, filter it ; adaptation of Python module of SyncPy. To 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.

## **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>



## 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 theoretical 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	Phd Student Social signal, 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	Prof Social signal, supervising the project, and the SyncPy library
David Cohen	Prof of Psychiatry, Original idea, supervising the project
Catherine Saint Georges	Psychiatrist, knowledge in interpreting synchrony scores in psychiatry
David Reversat	Informarician, Debugging of SyncPy
Michel Spodenkiewicz	Psychiatrist, Collection of the database (videos from INCANT and Monrado studies), direct supervision
Giovanna Varni	Post doc researcher Social signal, Development of the SyncPy library

The Monrado (Dr Coady Vulliez, Mrs Monika Szymanska) and Incant (Dr Phan, Mrs Bastard) 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.

## 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, with very different practices and level of evaluation. Contrary to other parts of medicine, psychotherapy and consequently psychiatry remains organized as schools which hardly understands and 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 operationalizing precisely the targets, Cognitive and Behavioral Therapies are the most evidence based psychotherapy which is caused in great part by a lack of experimental work in other schools of psychotherapy (Canceli et al., 2004 ; Layard & Clark, 2014).

Paradoxically, psychotherapies methods seem 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 patients to have a remission impact on a group of patient (Wampold cited by L'évaluation des psychothérapies Méthodes, résultats, réflexions, Xavier Briffault, Séminaire de Pierre-Henri Castel du 14 janvier 2009).

This is 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, we 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 (evidence based medicine approach) to compare it with another treatment like a previous classical treatment or a placebo. For instance, 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, it is costly because it takes time 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 has relaxation techniques, Cognitive and Behavioral Therapies 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) seems to be a very important part of the efficacy of psychotherapy (Horvath & Dianne, 1991 ; Martin, Garske, & Katherine, 2000 ; Tschacher, Haken, & Kyselo, 2015). Even it is difficult to operationalize, some psychometric tools were developed like the Working Alliance Inventory questionnaire (Corbière, Bisson, Lauzon, & Ricard, 2006). This leads 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. This consensus is challenged by some behaviorist that state that this determinant is only important when psychotherapist are not well trained and in poorly designed studies (Layard & Clark, 2014). This is the reflect too that the numbers of studies on psychotherapies outside Cognitive Behavioral therapies is very low.

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, 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 which if much used in first steps Cognitive Behavioral Therapies 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 (like : If I understand well, you told me that..., isn't it ?) especially the change discourse.
- Sum up
- Avoiding direct advices but stimulating it by questions
- Accept the ambivalence in a change
- Avoid judgement
- Searching for small and positive steps toward change
- Reinforcing any positive step
- Define and raising the values of the patient and not the therapist values
- Giving advices only when it is requested by the patient

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 developed by Beauvois and Joule after the work on cognitive dissonance by Festinger. This theory was developed and validated in control people without any mental illnesses but could be very interesting in psychiatric patients. 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.

We can see that even this very operationalized psychotherapy let place for a lot of improvisation and adaptation to a precise. Consequently, it needs a lot of practice and training beyond the theoretical knowledge. 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 "glace sans teint". An other way

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. However, it can be very difficult and fastidious.

In many parts of medicine, we don't directly evaluate the desired outcome of a therapy (like the mortality) because it happens to 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 and mortality). Consequently, it is very convenient. 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 versus classical one.

It is possible to evaluate the efficacy of a psychotherapy via two 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 cognitive behavioral 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 cognitive behavioral 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.

### **Attention**

Dehaene (Neuroeducation, Curiosity, Sleep)

### **Cognitive dissonance and engagement theory**

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, 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 another 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).

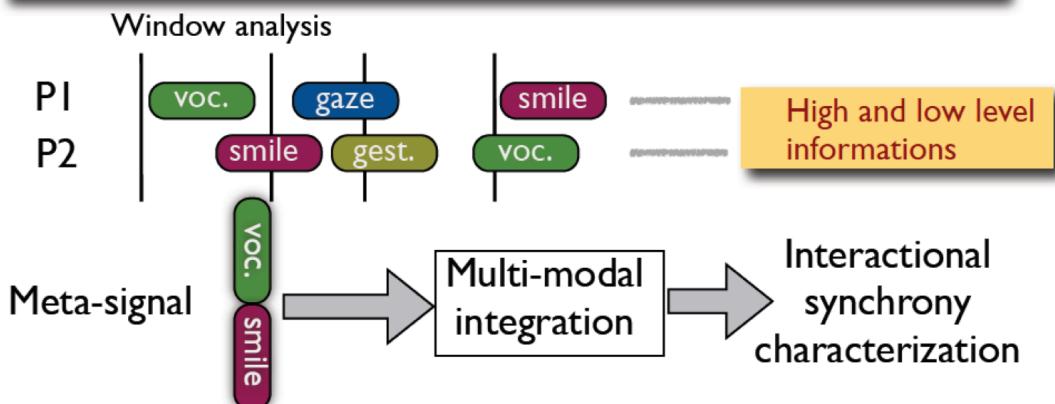
Behavior matching [21]; mirroring; mimicry [22], [23], [24]; congruence and the chameleon effect [25] are related to convergence

Synchrony is related to the adaptation of one individual to the rhythms and movements

of the interaction partner (Condon & W. Ogston, 1967). For Delaherche, "synchrony is the dynamic and reciprocal adaptation of the temporal structure of behaviors between interactive partners" (Delaherche 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).

### 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.

# **Attachment and psychotherapy**

---

TODO

## **Open science, Open data, Confidentiality**

---

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, we tried to develop an open science 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.

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 in challenges that enable the development of the best classifier for instance.

We will describe in a IMRAD organization work on 2 databases : INCANT and MONRADO.

## **INCANT Study**

---

### **Introduction**

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 like topics that needs to be addressed, attitude of the therapist.

## Material and methods

### Study protocol

#### Videos

- We collected a 252.44Gb database of 277 Videos, 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)
- [Mou](#) : Markdown editor
- [Git](#)
  - [GitHub website](#) public version control repository

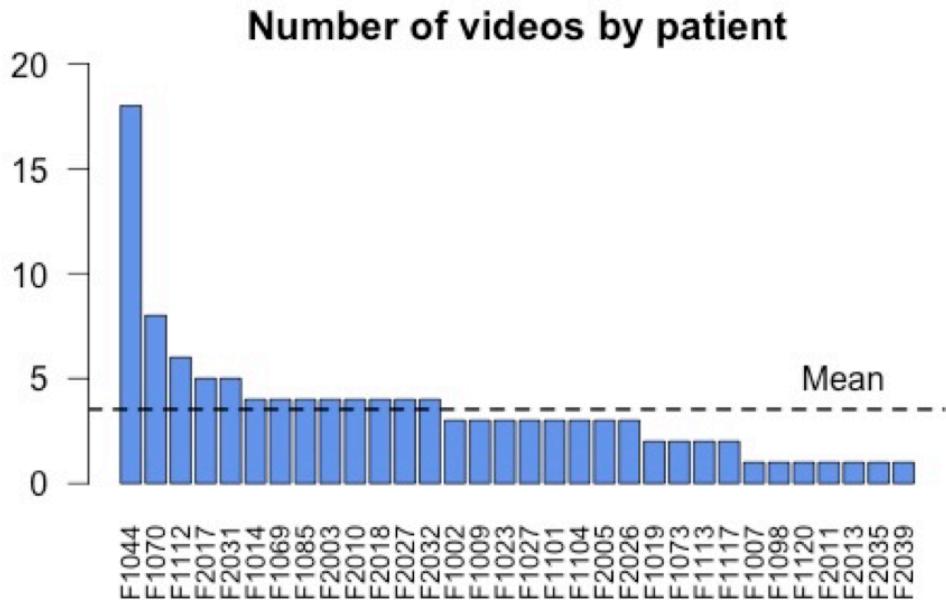
## Psychometric data

- 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: F1044 subject, his father, mother and therapist

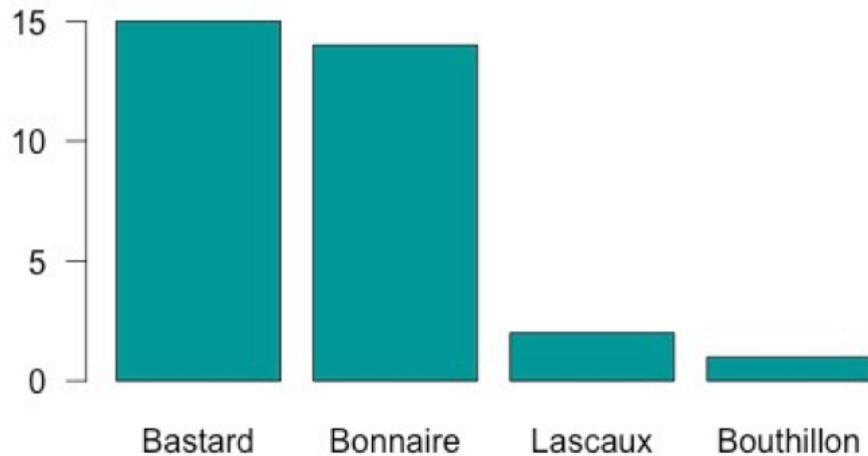
We decided to make a pilot study on the most complete subject : F1044 since we have 18 videos with him or his family.



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.

## Number of Patients by therapist



- 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).

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

## Description of the database

## Nomenclature

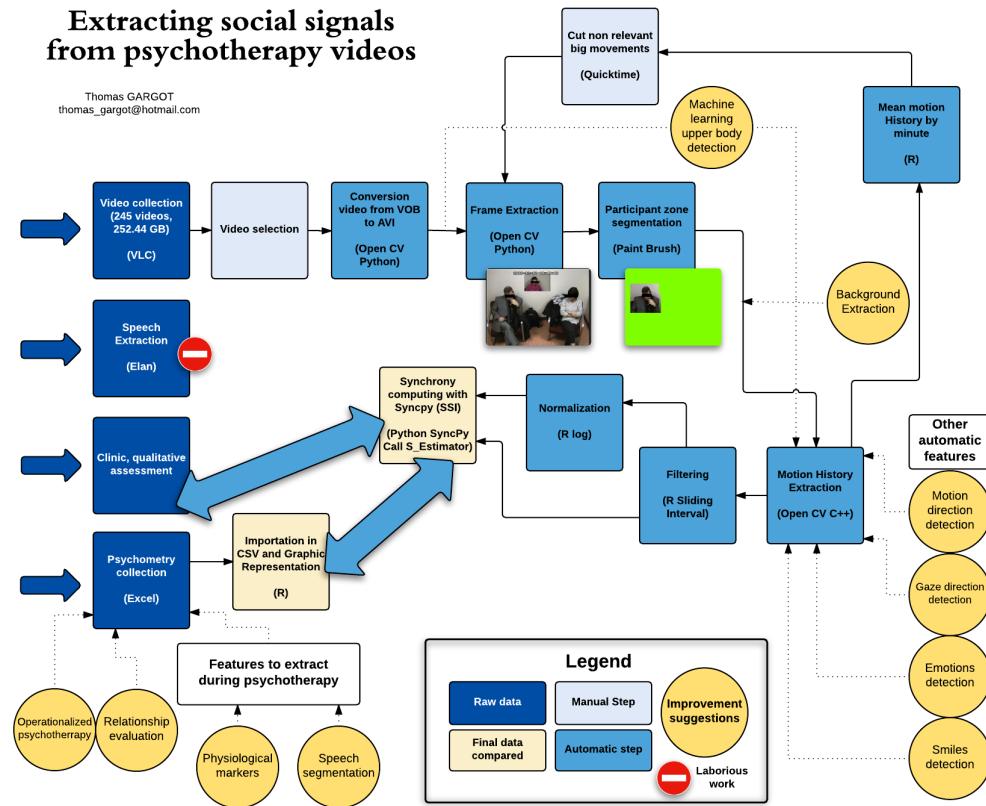
*F1044* is the name of the subject studied (called *patient*). When referring to several persons, variables take the name of the shorter form *pa* for *patient*.

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 family 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 subdivided after that with numbers (eg *F1044C*).

## Summary figure of technical steps



## Video Configurations

### Verbal signal extraction

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. This waveform helped to identify the zones of speaking vs no speaking.

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.

Consequently, this process was very laborious. It was possible to annotate one hour of video with a whole week work.

We tried to find ways to make source separation automatic even if it couldn't be perfect.

We found the [FASST software](#) that is very powerful to separate sounds coming from different music instruments. However, we contacted the author of this software and this material is too difficult to separate at the moment since two voices are not so much different from two different music instruments.

### **Verbal signal extraction discussion**

This separation could be made during the record of the video with several mikes that could be small tie mikes attached to each participant.

It is often necessary to extract manually some features even because of the lack of automatic process or in case of developing, training and evaluating these tools.

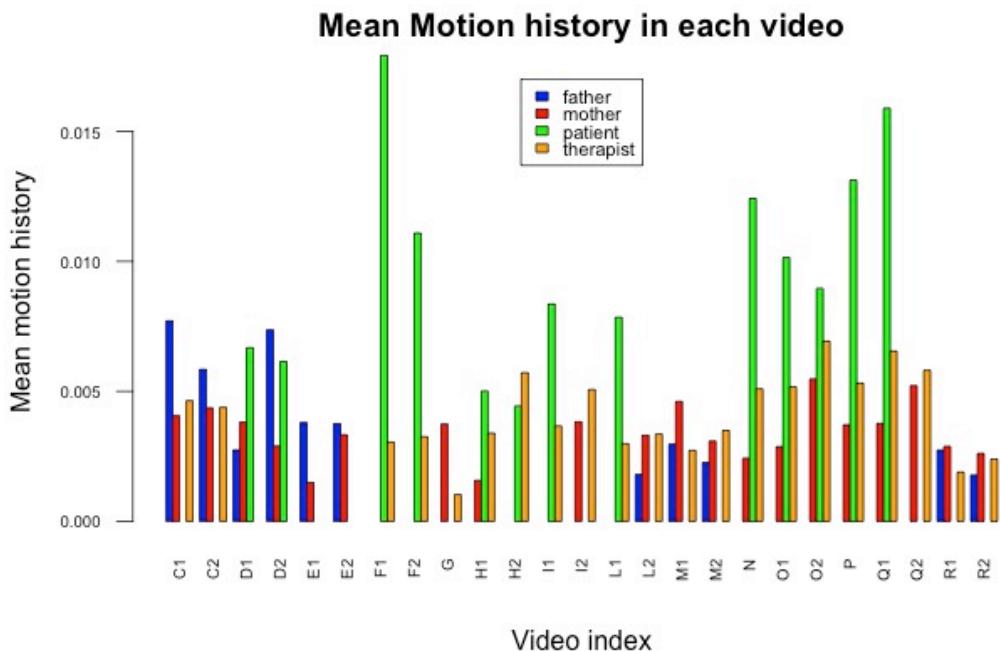
Consequently, we decided to give up this approach.

### **Spatial organisation**

In the first center, psychotherapy sessions consisted of filmed familial psychotherapies with 2 to 5 peoples organized in a circle. 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.

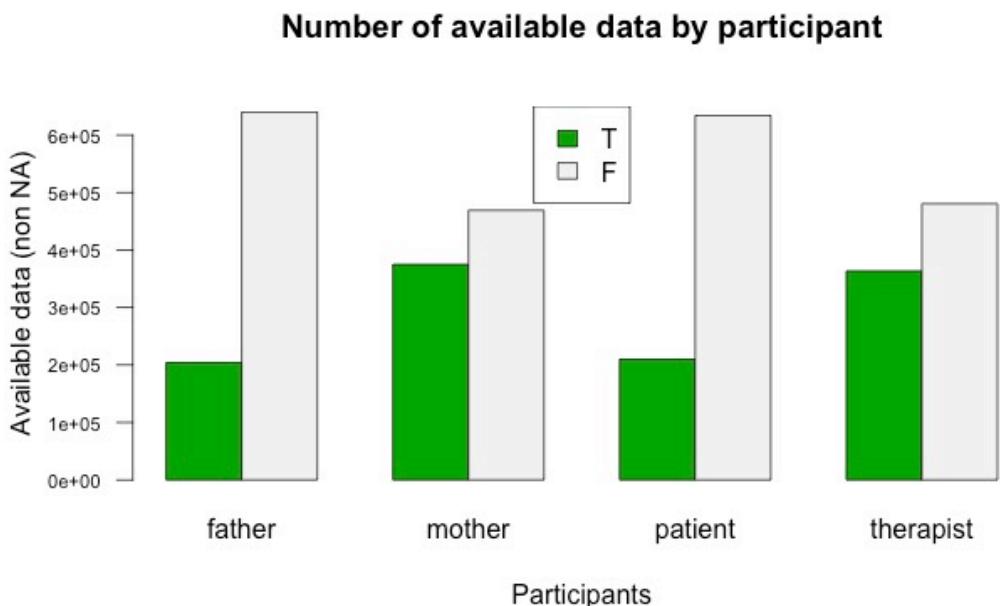
Sometimes, there is an overlap of different subject in the same place.





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).

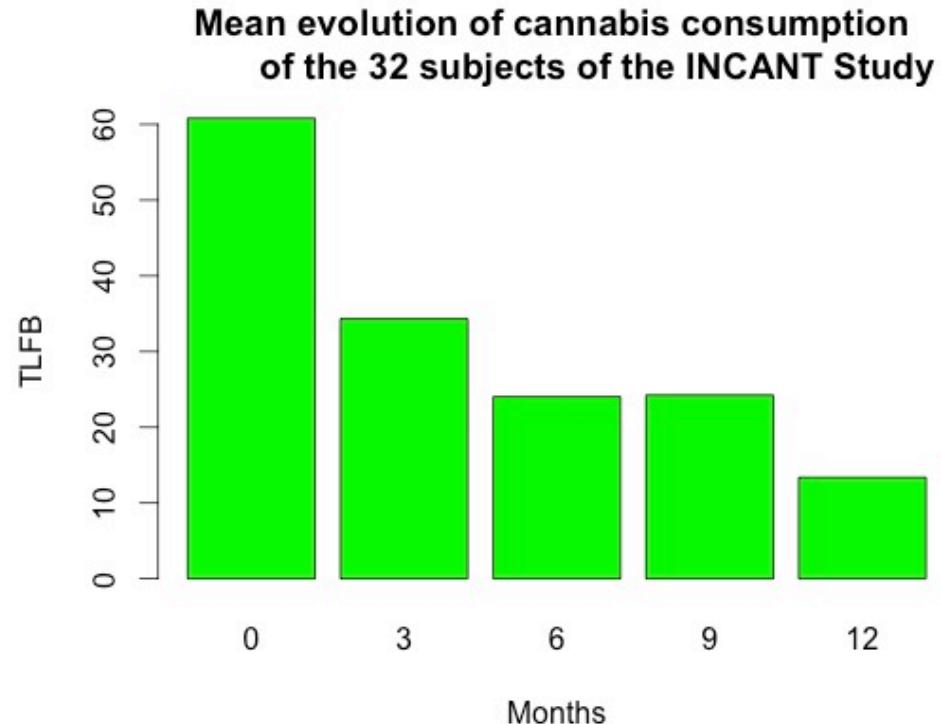
When we sum up the data we get, we can see that the mother and the therapist are the most present participants.



## Psychometric data

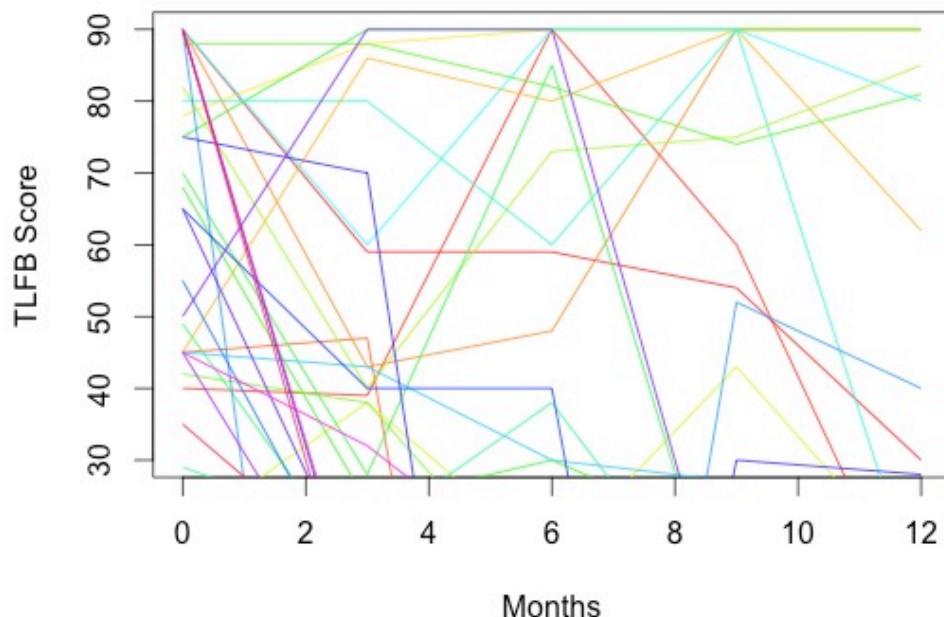
Files were collected in [xls](#) format. They were imported via the [XLSToCSVConvertor.R](#) R script. It returns a `dataINCANT.csv` csv file and a [dataCannabis.csv](#) file.

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.

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

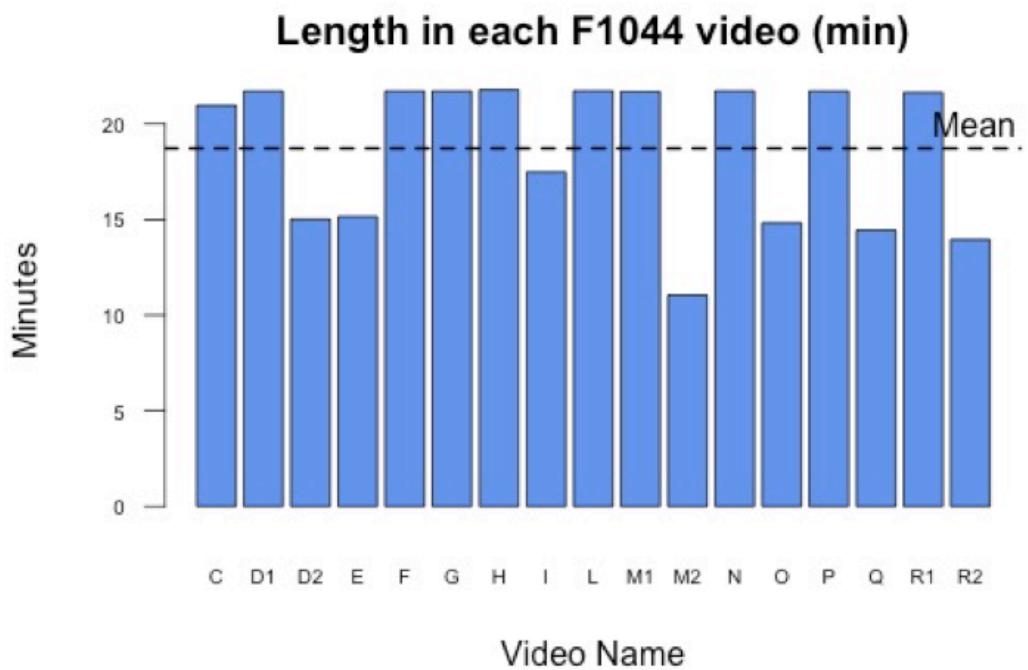


All the evolutions of this scores are plotted [here](#).

### **Time organization**

Video length could be 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 <a href="#">challenge sessions</a> 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](#).

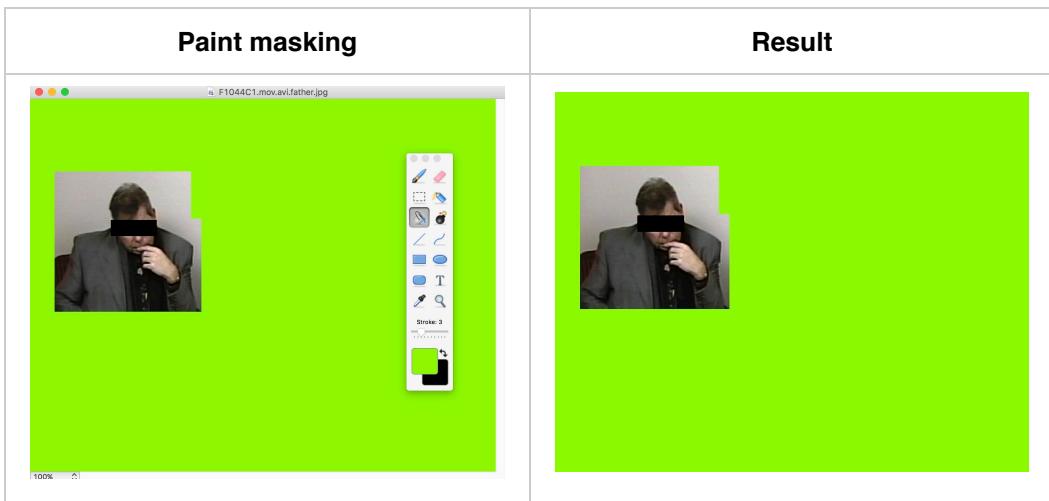
Here is an example of the father (eyes blurred to keep anonymity).



Idea	
	It was suggested to take a mean image of the video to be more precise instead of taking a video from the beginning.
	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 <a href="#">kinect</a> . A lot of other features could be extracted but the volume of this data can be very heavy several megabytes by minute of video. This is much more difficult to record for psychotherapist but would be much more relevant than raw video alone.
Beware	
	Notice the date that was always masked after during the step 2.
	We can change the value of the <a href="#">median filter</a> .

## 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.



This operation is repeated for each participant on each video.

Beware	
<b>BEWARE</b>	It is necessary to always use the same background color, for instance <a href="#">RGB green</a> . (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 <i>F1044C1.avi.father.jpg</i> )
Idea	<p><a href="#">Deep learning software</a> or machine learning used for categorization/detection of people could make this process completely automatic.</p> 

### 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 :

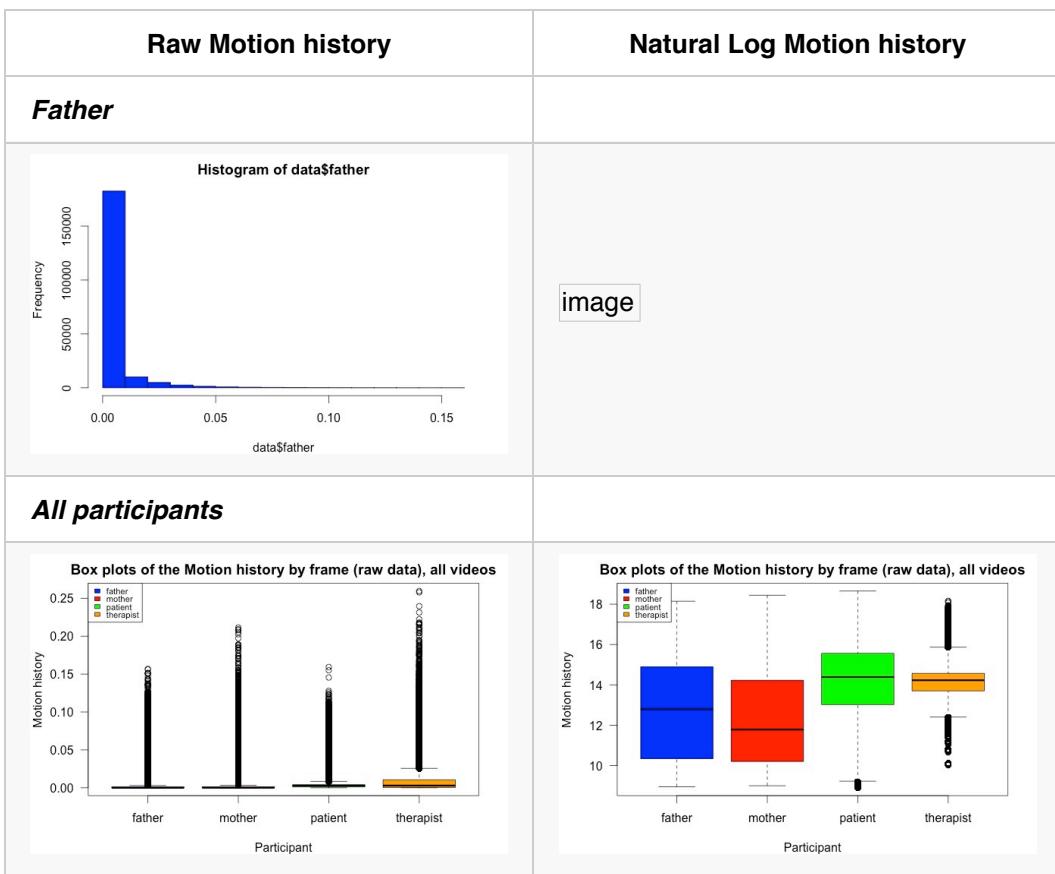
- **frame** : number of the frame (original rate of 25 frames by second)
- **father** : motion history of the father frame by frame. It ranges from 0 (same frame than the previous frame, no movement) to 1 (every pixel change).
- **mother**, : idem for mother
- **patient** : idem for patient
- **therapist** : 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 raw data (1st column), 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.

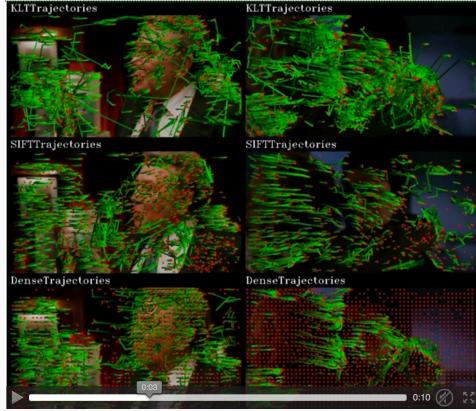
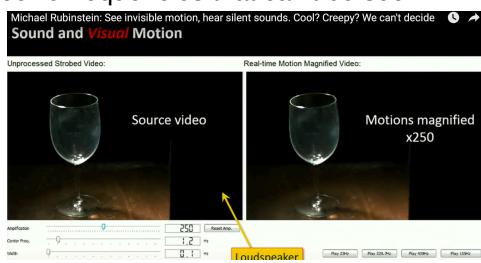
To normalize the distribution to compute synchrony scores on it, we made the natural logarithm. It produces negative numbers. SyncPy can't compute negatives scores, they are so shifted to positive values with an arbitrary value of 20 to avoid to keep extreme negative values.

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 an arbitrary value to this data (eg, the minimum value, they are over represented).

This normalized log data is shown on column 2.



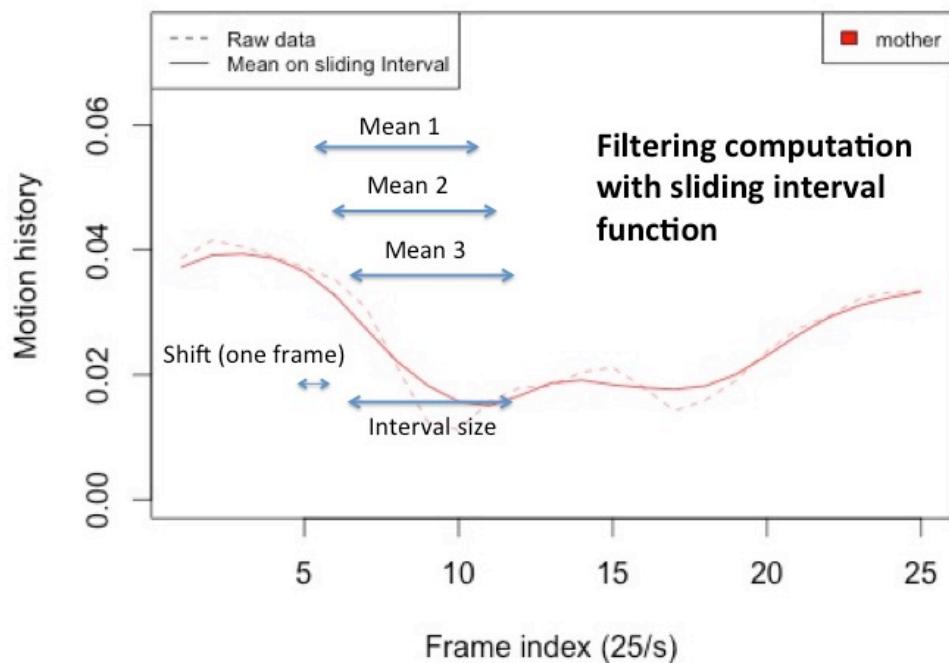
If we check all participants, we can see that the father and the mother motion history distribution is very similar. However, the therapist, which is always in a small window of the video, as a very different distribution. We have less signal on it. In some videos the patient is in this window, it explains,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

Idea	
	<ul style="list-style-type: none"> <li>- It could be possible to extract <a href="#">background</a> to improve the quality of the motion history. See for instance <a href="#">backgroundExtractor.py</a> with openCV <a href="#">BGS library</a>.</li> </ul>
	<ul style="list-style-type: none"> <li>- It could be interesting to evaluate the trajectories of the people. If they are going toward the same direction or not. For that, the <a href="#">optic flow modules of open CV</a> or the <a href="#">Dense Trajectories Video Description</a></li> </ul>  <p>could help.</p>
	<p>Tiny changes can't be always obvious. <a href="#">Some softwares</a> can amplify a lot very tiny changes in specific frequencies that can't be seen normally with naked eyes.</p> 

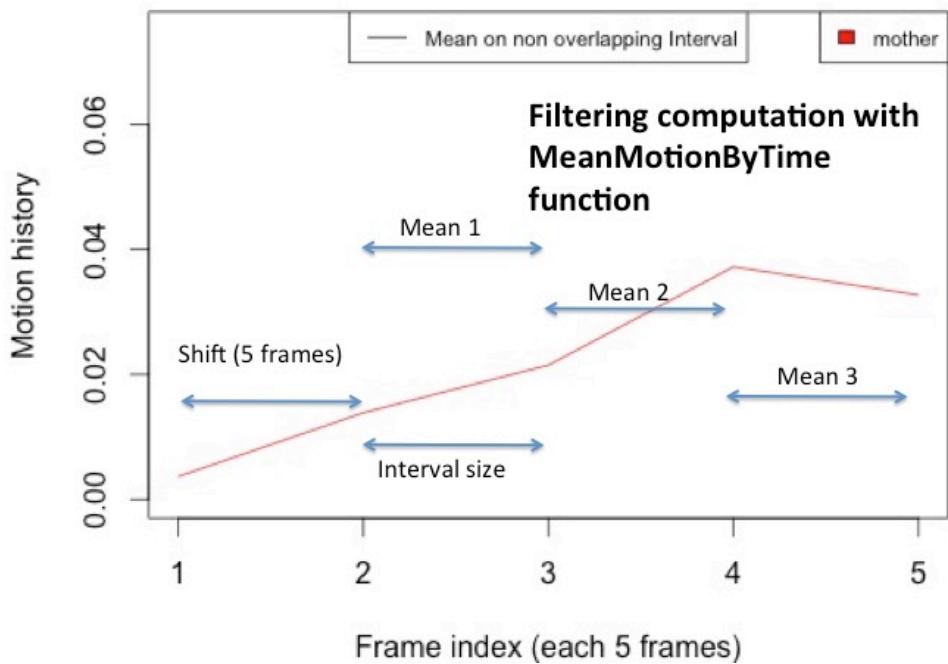
#### 4. Motion history filtered with slidingInterval function

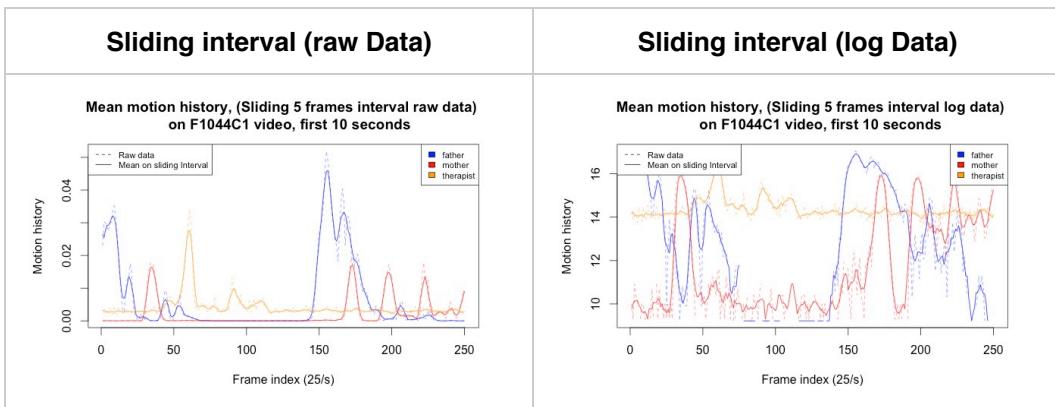
Motion history raw data is filters with a [R](#) function on [Rmarkdown script analysis](#). It generates CSV file of the form "F1044C.VOB.slidedata.csv". See files [here](#).

**Mean motion history (Sliding 5 frames interval)  
on 00034 video, 2nd second**



**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
- **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 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).

## B. Speech extraction

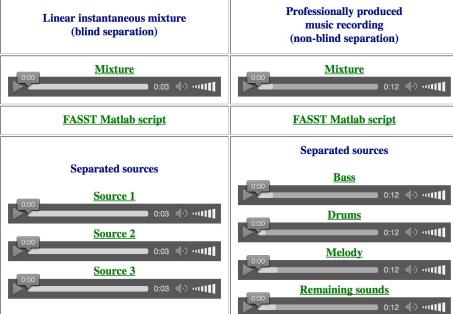
### 1. Manually

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

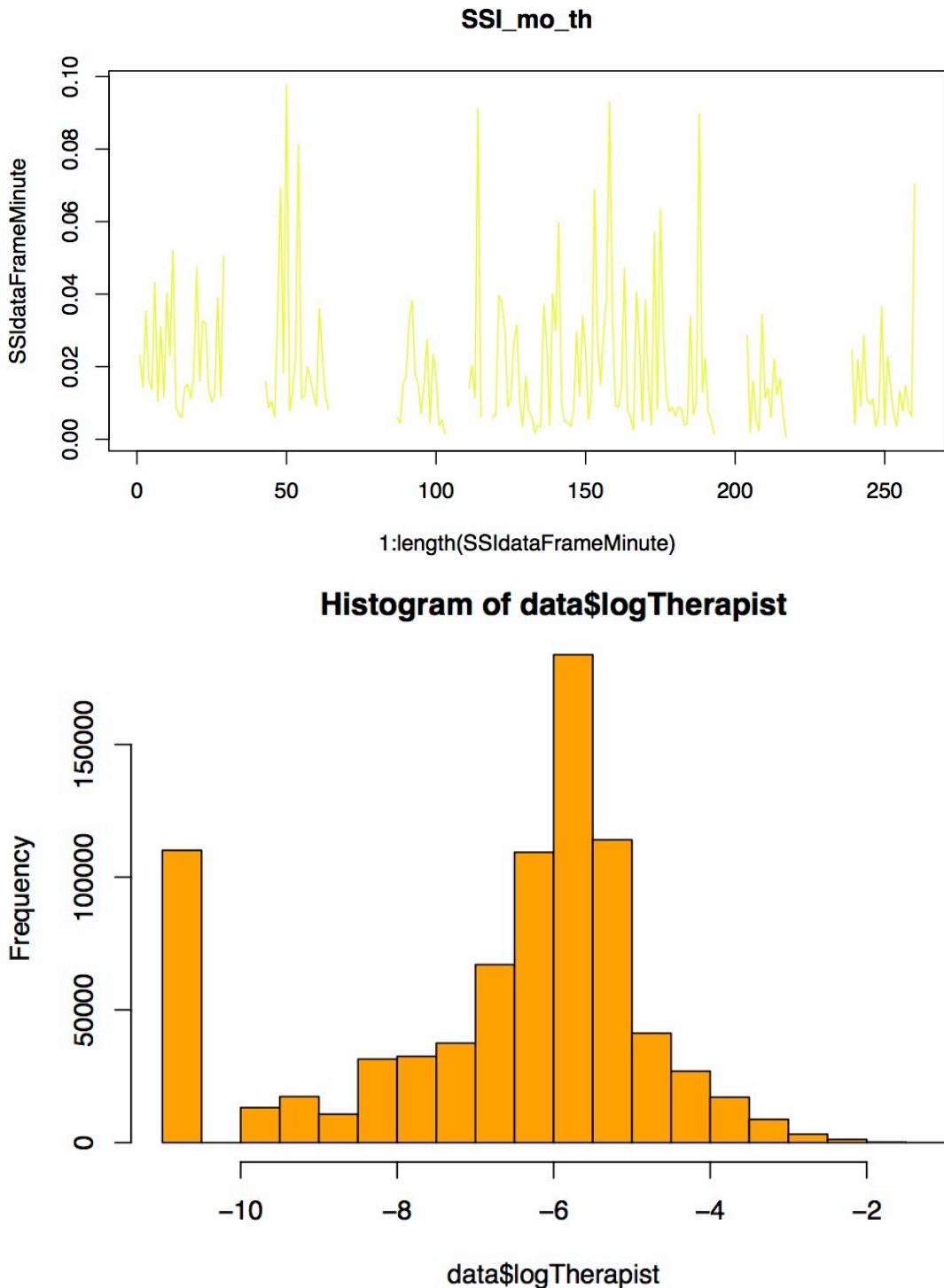
Idea	
	<p>Some softwares could be used to make this process easier. The <a href="#">FASST software</a> get very good results to segment a music sample in very different instruments. However, different voices are too close to distinguish them easily automatically.</p> <p><a href="#">Source separation examples</a></p>  <p>The screenshot shows a comparison between a 'Linear instantaneous mixture (blind separation)' and a 'Professionally produced music recording (non-blind separation)'. It illustrates how the FASST Matlab script can separate a complex audio mixture into distinct musical instruments like Bass, Drums, Melody, and Remaining sounds.</p>

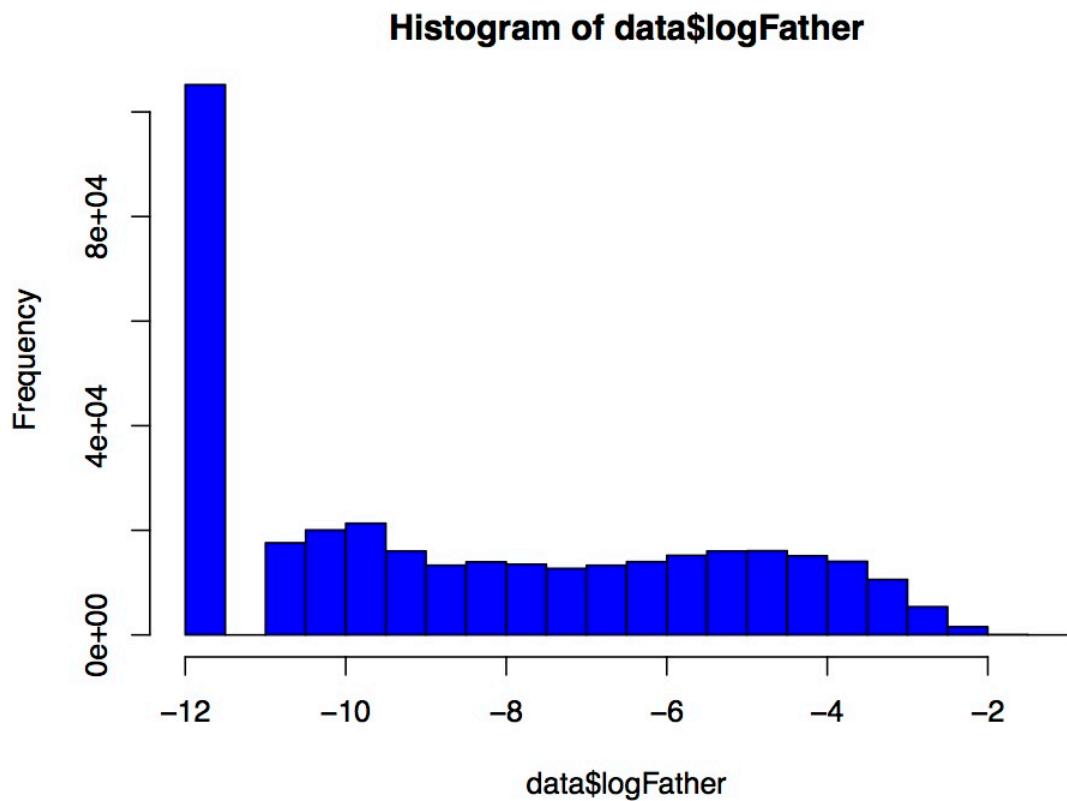
We can see that the evolution of cannabis use by F1044 is not straightforward. It is quite high at the beginning, decreases and then increases. Consequently, we decided to select 4 subjects with a clear decrease of cannabis consumption.

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

All of them with the same therapist Mrs Bastard

## Results





## Discussion

### 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

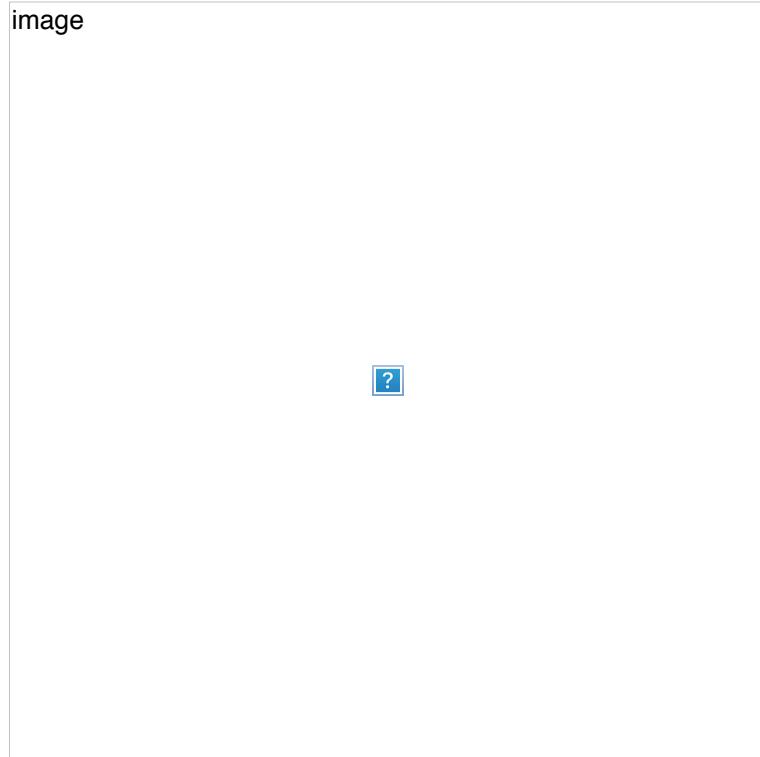
#### Experimental set-up

It is however not very defined in the articles. Consequently, it was difficult to contrast synchrony between two well defined periods. We could suggest to have a much more

defined psychotherapy and even before studying a psychotherapy to study specific situations : open vs closed questions,

Well defined Motivtional interviewing CBT not enough evidence based

### 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.

image



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.

image



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

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.

Furthermore, in next protocol, it would be considered to use either a kinect, either, to keep always the same orientation of the subject toward the camera.

### Technical discussions

It seems that it would be very important to have several devices (camera, kinect, mikes). 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.

Multi person : not the majority of psychotherapy more often filmed for supervision and educational use.

shared database and problems of confidentiality

## MONRADO study

---

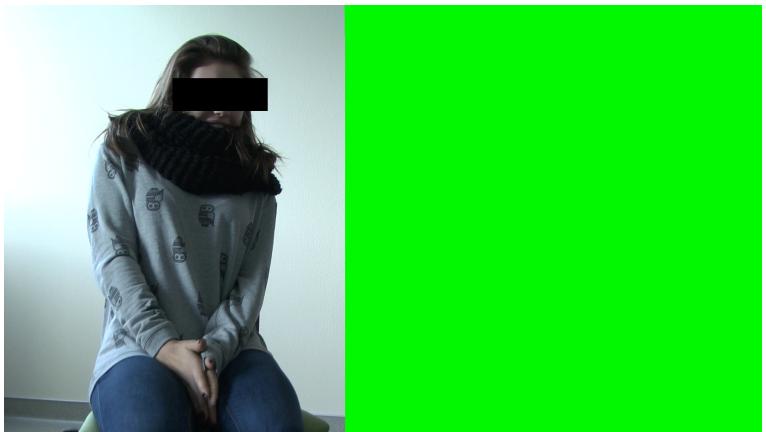
### Aim of the study

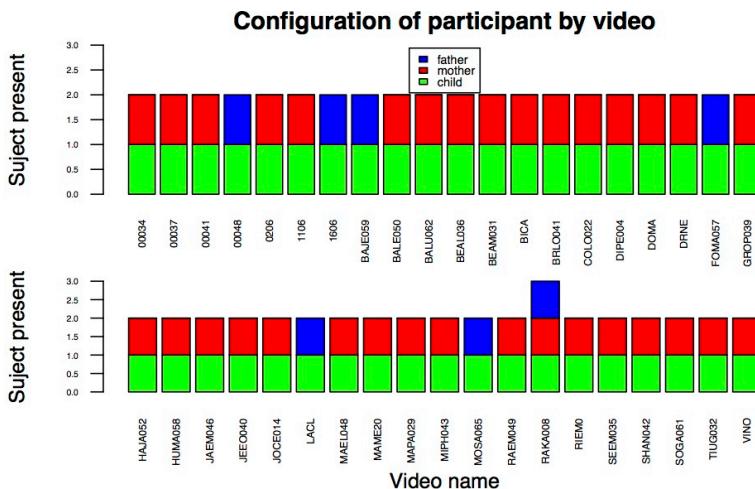
The aim of the study was to evaluate the response to conflict in a dyad of a parent and a child. There were two phases. After the experimenter left the room, the dyad spoke about a casual topics. After around 5 minutes, the experimenter came back in the room and asked the child to speak about a conflictual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

Consequently, from these videos, we can extract easily two phases : 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 no overlap between the subjects. There isn't embedded subjects. There is no date embedded.





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

## 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

## 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) seems to be user-friendly, transparent and reliable measure?

It offers the adolescent 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. Report of a voluntary sector and university partnership ('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/>

Attachement 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

### 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

M.P. Marchand et G. Loas (1994).

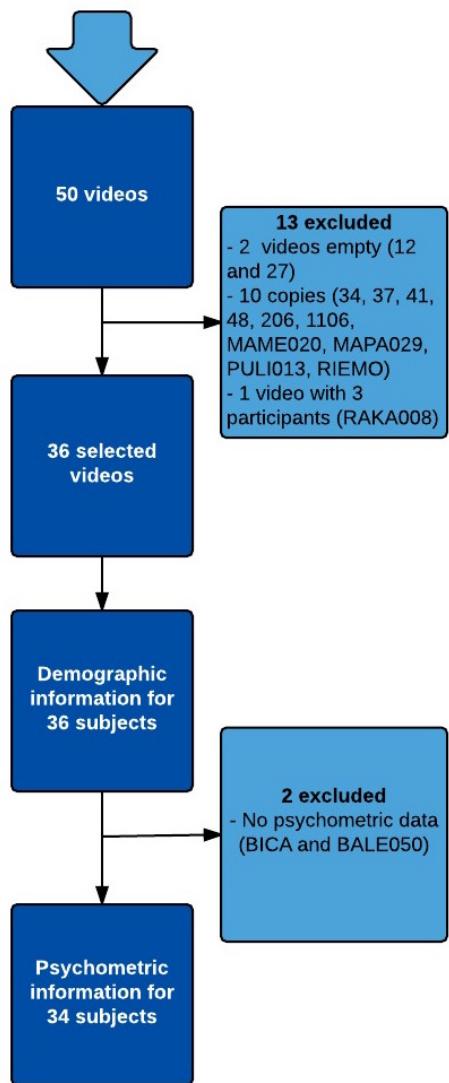
### STAIYA Score

The STAIYA or State-Trait Anxiety Inventory of Spielberger is an auto-evaluation of Anxiety. It is one of the most used Anxiety questionnaire. It dissociate the anxiety at a precise moment (state) and the habitual anxiety of the subject (trait). It was evaluated on more than 5000 subjects.

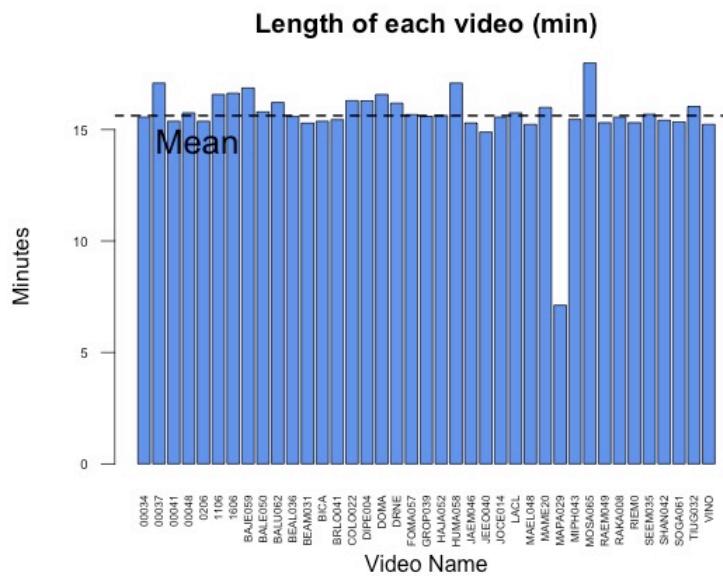
### Description of the database

- 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.

### Selection of relevant videos

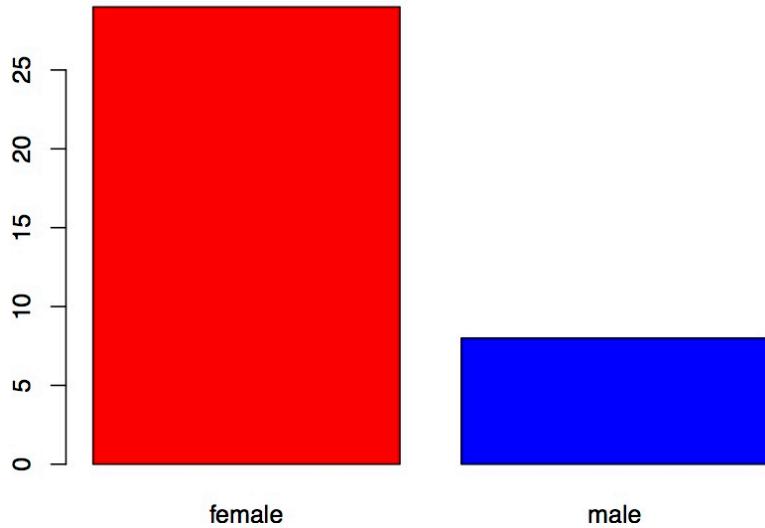


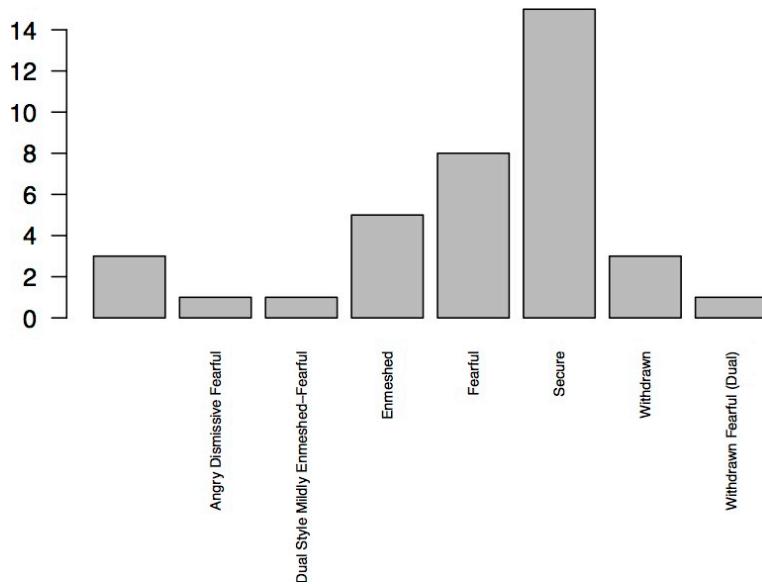
#### Length of the videos



We can see that the videos are very similar.

### Sex repartition





### Technical problems

None There was only one video which never moved. There wasn't any date. There was very few overlap of the subjects.

### 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
  - oxytocin
- Cardiac Frequency : direct or indirect on videos with systems like [Lambda](#) or [webcam pulse detector](#)
- Annotation 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 very old since it was used in [psychoanalysis by C. G. Jung](#).
- Motion capture Optitrack
- [Ethome](#) and special suits

- Kinect : skeleton
- Source separation source : FASST
- Facial / Upper body recognition to segment images
- precise camera ? for each person ?

background subtraction gaze direction optic flow and trajectories

## Discussion

---

### Experimental discussion

It would be interesting to Assertiveness Motivational interviewing provocation relaxation exposition dream analysis speaking of its own or other emotions Gaze Emotions classifier (Smiles) Hupont, I., Baldassarri, S., & Cerezo, E. (2013). Facial emotional classification: from a discrete perspective to a continuous emotional space. Pattern Analysis and Applications, 16(1), 41-54.

Head Pose (Head direction) De la Torre, F., Chu, W. S., Xiong, X., Vicente, F., Ding, X., & Cohn, J. (2015, May). Intraface. In Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on (Vol. 1, pp. 1-8). IEEE.

Analysis discussion Modeling, Machine learning

Shuffle : non random synchrony signal