Aim and Hypothesis

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

Here is the analysis of the F1044 subject, his familly (2 parents) and the therapist. Is there synchrony signals between himself, his parents and the therapist ?

Is there synchrony signal between his parents and the therapist ? Could this synchrony signal predict outcomes of the psychotherapy ? See the full pre-registration.

We used this data from a large european psychotherapy study since they come from almost real life practice of psychotherapy. This INCANT study. study aimed to evaluate the efficacy of the MultiDimensioinal Family Therapy for cannabis use disorders in adolescents.

Nomenclature

F1044 is the name of the subject studied (called pa for patient). He has a mother (mo) a father (fa) helped by a therapist (th). When a variable is referring to several participants, it is organised in alphabetical order separated by underscores, eg. SSI_fa_mo refers to the synchrony index (SSI) between the father and the mother. SSI_mo_fa doesn't exist. This familly had several consultations with the psychotherapist. Some of them were video recorded. These videos are names with the name of the subject + an index letter. They can subdivised after that with numbers (eg F1044C).

Steps

Raw data

It is the videos in AVI format in the form F1044C1.avi, with a rate of 25 frames by second, from the 2 French centers of the INCANT study.

This raw data is compared with psychometric data.

This data was made with public funds, however, the access of it was difficult, moreover it wasn't collected for this goal.

- In a lot of other domains of fundamental algorithm development, computer labs compete on public databases to improve the detection rates of some features for instance. This work are presented during challenge sessions in congresses.

- 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 a python script from the beginning of the video to make a mask.

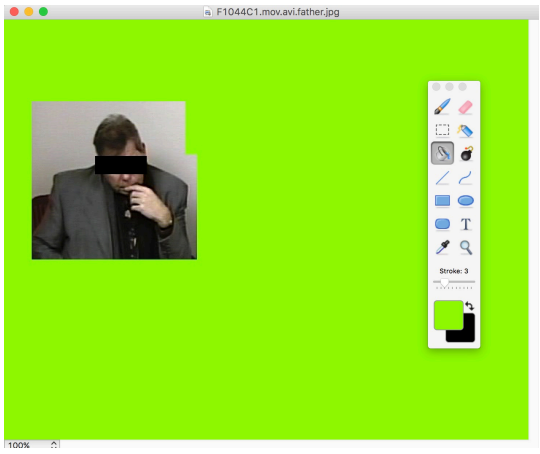
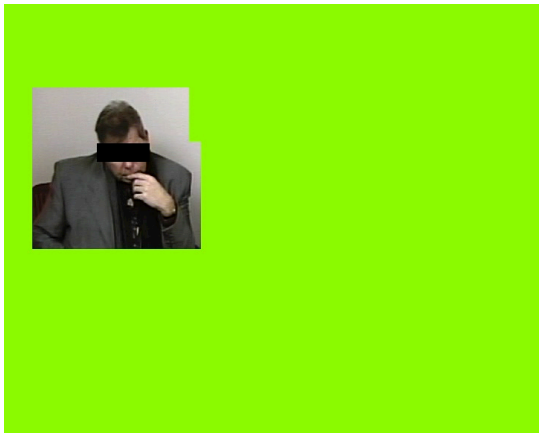
Here is an example of the father (eyes blurred to keep anonymity).

It was suggested to take a mean image of the video to be more precise instead of taking a video from the beginning.

Notice the date that was always masked after during the step 3.

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 paint brush. Take only the upper part of the body to compare them since you can see only the upper part of the therapist.

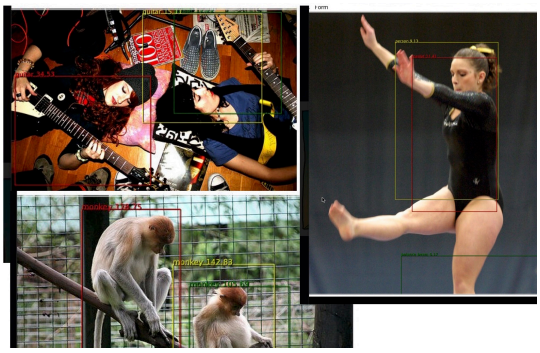
Paint masking	Result
	

This operation is repeated for each participant on each video.

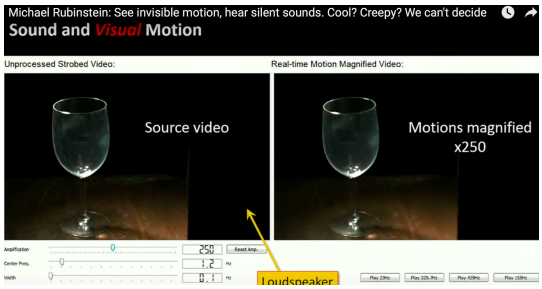
It is necessary to always use the same background color

Participants can move during the video (change their seats, leave the room). It would be preferable to anticipate it. Mean Motion history by minute can help us to detect big changes or disappearance of a participant.

- Deep learning software used for categorization/detection of people could make this process completely automatic.



- 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"



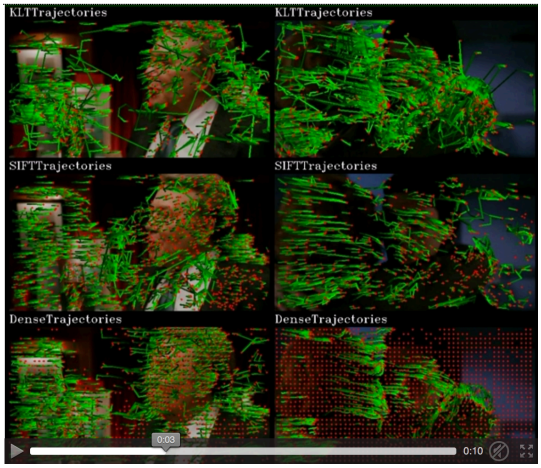
3. Motion history extracted

It is extracted with the video and the mask 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
- **mother**, : idem for mother
- **patient** : idem for patient
- **therapist** : idem for therapist
- **file** : name of the file in the form F1044C.VOB
- **NA** corresponding to absent subject or bad quality of first frames at the beginning or end of the videos

- It could be possible to extract background to improve the quality of the motion history. See for instance BGS library.
- It could be interesting to evaluate the trajectories of the people. Are they going toward the same direction or not. For that, the Dense Trajectories Video Description could help.



4. Motion history filtered with slindingInterval function

CSV file of the form "F1044C.VOB.slideddata.csv". See files here.

- **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)
- **slidedMother** : idem for mother
- **slidedTherapist** : idem for therapist
- **slidedPatient** : idem for patient
- **NA** corresponding to absent subject

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

- 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

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

It returns synchrony CSV files that can be found here

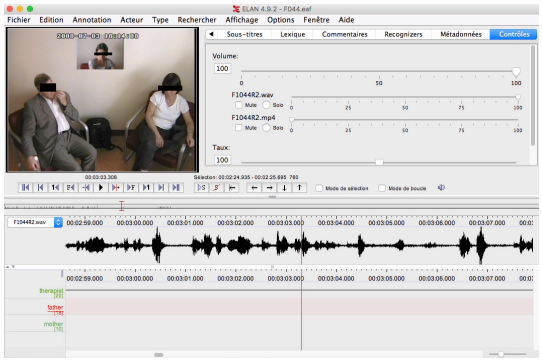
- **Interval**
- **Time_min**
- **video** : name of the video (eg F1044C.VOB)
- **SSI_fa_mo** : synchrony index between the filtered motion history of the father and mother
- **SSI_fa_mo_th** : idem between father, mother and therapist
- **SSI_fa_th** : idem between father and therapist
- **SSI_mo_th** : 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 was gave up).

2. Automatically

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

Source separation examples

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