REFLETS: Rétroaction **É**motionnelle **F**aciale et **L**inguistique**,** et **É**tats de stress post-**T**raumatique**s**

Table of contents

Project summary

La théorie de la rétroaction émotionnelle périphérique – le fait que nos expériences émotionnelles soient sous l'influence rétroactive de nos propres expressions - est un sujet ouvert en psychologie depuis les écrits de William James au 19e siècle. Que le fait de sourire ou de froncer les sourcils puisse avoir un effet automatique dans l'expérience émotionnelle d'une personne ouvre d'importantes pistes thérapeutiques pour les troubles psychiatriques de l'émotion. Malheureusement, cette idée est confrontée à des défis théoriques et méthodologiques majeurs: en Juillet 2016, 17 laboratoires n'ont pas réussi à reproduire l'effet affiché dans l'un des articles fondateurs du domaine, en dépit du fait que l'étude originale ait été citée plus de 1500 fois. Le Projet REFLETS vise à remédier à cette situation en proposant une nouvelle technologie de transformation audio/visuelle permettant de canaliser le potentiel du mécanisme psychologique de rétroaction émotionnelle faciale et vocale pour une application clinique dans le domaine du stress post-traumatique (PTSD).

The theory of peripheral emotional feedback – that our emotional experiences are under the retroactive influence of our own expressions – has been an ongoing subject of debate in psychology since William James. On the one hand, the fact that putting on a smile or a frown may have an implicit, automatic effect in one's emotional experience holds tremendous potential for clinical remediation in psychiatric disorders. On the other hand, the idea faces tremendous theoretical and methodological challenges: in July 2016, a 17-laboratory Registered Replication Report failed to reproduce the effect found by one of the seminal paper of the field, despite that original study being cited more than 1500 times. Project REFLETS aims to address this situation head-on, building new health technology able to channel the psychological mechanism of facial and vocal emotional feedback for clinical application to post-traumatic stress disorders (PTSD).

Partner	***Name***	***First name***	***Position***	***P.month***	***Role***
IETR	*SOLADIE*	*Catherine*	*Associate professor*	*18*	*Coordination/ cognitive psychological experiments*
IETR	*SEGUIER*	*Renaud*	*Professeur*	*8*	*Valorization*
IETR	*RICHARD*	*Pierre-Yves*	*Professeur*	*4*	*Miror calibration*
IETR	*xxx*	*xxx*	*PhD*	*36*	*Emotional software*
IETR	*xxx*	*xxx*	*Post-doc*	*7*	*Emotional software / cognitive experiments*
IRCAM	AUCOUTURIER	*Jean-Julien*	*CR1 CNRS*	*15*	*Cognitive psychology experiments (SP 4)*
IRCAM	LIUNI	*Marco*	*CR IRCAM*	*4.3*	*Audio smile transformations (SP 3)*
IRCAM	xxx	xxx	CR IRCAM	5	*Audio smile transformations (SP 3)*
IRCAM	xxx	*xxx*	*PhD*	*36*	*Cognitive psychology experiments (SP4)*
Dynamixyz	AUBAULT	*Olivier*	*CTO*	*6*	*Coordination*
Dynamixyz	STOIBER	*Nicolas*	*Research Head*	*20*	*Camera calibration /emotional software*
Dynamixyz	BARRIELLE	*Vincent*	*Research Ing.*	*28*	*Camera calibration*
HumanEvo	CLAUSSE	*Jérémie*	*Président Directeur Général*	*12*	*Well-being applications (SP4)*
Chanel	PORCHERON	*Aurélie*	*Chercheur*		
Cognac-G	BUFFAT	*Stéphane*	*Chercheur*	*8*	*Health clinical applications (SP4)*
Cognac-G	VIDAL	*Pierre-Paul*	*Directeur de Recherche CNRS*	*2*	*Behavior monitoring*
Cognac-G	RICARD	*Damien*	*Professeur agrégé du Val de Grâce, neurologue*	2	*Health clinical applications (SP4)*
Cognac-G	BOMPAIRE	*Flavie*	*Neurologue*	*2*	*Health clinical applications (SP4)*
Cognac-G	MOREAU	*Albane*	*Ingénieure Biomédicale*	*2*	*Health clinical applications (SP4)*
Cognac-G	xxx	*xxx*	*Post-doc*	*12*	*Health clinical applications (SP4)*

I. Proposal's context, positioning and objective(s)

I.1. Objectives and scientific hypotheses

The theory of peripheral emotional feedback – that our emotional experiences are under the retroactive influence of our own expressions – has been an ongoing subject of debate in psychology since William James (1884, Niedenthal et al, 2010). On the one hand, the fact that putting on a smile or a frown may have an automatic effect in one's emotional experience holds **tremendous potential for clinical remediation** in psychiatric disorders (an estimated 40-75% of which are linked to problems of emotional regulation - Gross & Jazaieri, 2014). On the other hand, the idea faces **tremendous theoretical and methodological challenges**: in July 2016, a 17-laboratory Registered Replication Report (Wagenmakers et al, in press) failed to reproduce the effect found by one of the seminal paper of the field, Strack, Martin, and Stepper (1988), despite that original study being cited more than 1500 times (Google Scholar, Mar. 2017).

Project REFLETS aims to address this situation head-on, by building novel health technology able to channel the psychological mechanism of facial and vocal emotional feedback for clinical application to post-traumatic stress disorders (PTSD) as well as well-being applications in the general population. To do so, project REFLETS combines significant research strengths in the three fields of Computer Science (audio/visual signal processing; lead by a team ranking first in two international challenges in facial emotion recognition), Cognitive Psychology (affective and social neuroscience; lead by a team funded by an ERC grant), and Clinical Psychology (human factors and post-traumatic stress disorders; lead by the French Army Medical Institute), in addition to support for design and product development from one of the major player of the luxury industry (Chanel).

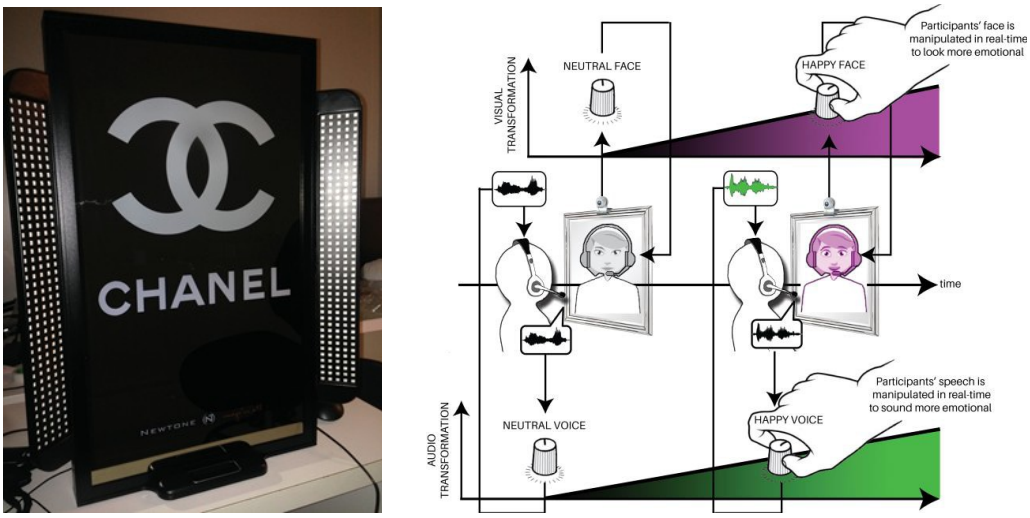

***Figure 1**. Prototype of computer-augmented mirror (left) and its proposed functionality (right): the observers' reflected face (captured by a set of cameras) is algorithmically and photo-realistically transformed to appear more smiling, and their speech (captured by a microphone) is re-synthesized in real-time to seem more happy and relaxed.*

The project's central technology is a computer-augmented mirror, in which observers both see and hear themselves in a gradually more positive way: without their knowing, their reflected face (captured by a set of cameras) is algorithmically and photo-realistically transformed to appear more smiling, and their speech (captured by a microphone) is re-synthesized to seem more happy and relaxed. Using the mirror, we expect that observers will come to believe the emotional tone of their transformed facial and vocal reflection as their own, and align their feelings with the transformation (see Aucouturier et al., PNAS 2016 for a recent proof of concept). Clinically, we propose to use this mechanism to remediate emotional numbing in post-traumatic stress disorder (PTSD) patients. We also propose to valorise the tool as a well-being device in the general population in the related context of professional burn-out and mild depressive episodes.

Project REFLETS will reach three major objectives, original with respect to each of its core disciplines:

Objective 1 (Computer science): ***Realistic, real-time, multimodal synthesis of emotional expressions.***

We will develop a software platform able to modify the audiovisual reflection of any given person, to make their face appear more smiling (i.e., increase the curvature of their mouth) and their voice sound brighter (i.e. shirt their formant frequencies in a way that simulates smiled speech). To this aim, we will develop a novel audiovisual deformation model, based on control points in the triangulated facial geometry (Soladié et al., 2013) and in the spectral envelope of the speech signal (Aucouturier et al., 2016). By objective 1, project REFLETS will go beyond the state-of-art in (1) photo- (and phono-) realism, by producing audiovisual expressions that the speakers themselves will believe are their own, and (2) in real-time performance, by doing these transformations with latencies that will not disrupt ongoing speech (a few tens of milliseconds).

Objective 2 (Cognitive science): ***Non-intrusive audiovisual peripheral emotional feedback.***

Using these novel audiovisual smile transformations, we will conduct cognitive psychology experiments in the vein of Aucouturier et al. (2016), to establish the methodological conditions of a strong emotional feedback effect amenable to the health and well-being applications of objective 3 (below). To do so, we will adapt the experimental paradigm of Aucouturier et al., 2016, in which participants read a text out loud while their voice was algorithmically modified to sound more happy or sad, and evaluate changes in the participants' emotional states using explicit (e.g. self-report before and after reading) and implicit measures (e.g. electrophysiology while reading). By objective 2, project REFLETS will go beyond the field's theoretical and methodological state-of-art by (1) eliminating demand effects (participants will simply watch their reflection in a mirror, without suspecting any manipulation) and (2) allowing unprecedented levels of experimental control and reproducibility, because algorithmic transformations will allow to explore all parameters (lip position, vocal features, etc.) of what makes a smile self-contagious.

Objective 3 (Health and well-being): ***Non-pharmaceutical remediation of emotional numbing for healthy adults and post-traumatic stress disorders (PTSD) patients.***

The project's societal arm is concerned with well-being applications in the non-pathological population, and with health applications in the clinical population of post-traumatic stress disorders (PTSD) patients. In these populations, we propose to target the clinical trait of emotional numbing or alexithymia (*loss of interest for previously pleasurable activities* – APA, 2013), which is a symptom of PTSD (Gross & Jazieri, 2014) but is also a risk factor in important well-being concerns such as occupational burnout (Mattila et al, 2007), mild depression (Honkalampi, 2000) and lack of empathy (Guttman & Laporte, 2002). To do so, we will integrate the audiovisual transformations developed as part of objective 1 into a custom hardware developed in collaboration with the project's industrial partner Chanel. Using this augmented mirror, we will conduct several longitudinal studies in which positive audiovisual transformations from the mirror are used to recondition once-pleasurable, now-bland daily experiences (e.g. going to the movies the day before), and to test the primary outcome of a more positive orientation of the patients/users towards the experience. Patients and controls for the clinical study will be included in the military and civilian PTSD population at Hopital d'Instruction des Armées Percy, while participants from the non-clinical population will be included at the former Hopital (now Ecole) Val-de-Grace Hospital. Both sites offer a state-of-art "smart flat" platform, in which patients and users will be able to interact with the device in ecological conditions, while being monitored with a variety of sensors useful for the study.

Final product: an augmented mirror device for health and well-being

The final product/outcome of project REFLETS is an audiovisual software and hardware platform, or "augmented mirror", able to improving its users' emotional attitudes towards past and future experiences by automatically transforming their reflection in the mirror. The hardware part of the platform will be developed by the Biology and Women's Beauty R&D of industrial partner Chanel, based on an existing prototype to be improved with several cameras and microphone to allow audiovisual transformations. This augmented emotional mirror is an ambitious example of new, interdisciplinary health technology, which incorporates both original computer science algorithms (the audiovisual emotional transformation developed as part of objective 1), novel psychological insights (the parameters and configurations, established as part of objective 2, to guarantee meaningful psychological effects to the participant) and is thoroughly validated with both clinical and user studies (objective 3).

I.2. Originality and relevance in relation to the state of the art

I.2.1 Audiovisual Emotional transformations

Smiles (and other outward signs of enjoyment in the face and voice) are a fundamental element of the human expressive repertoire. We recognize smiles regardless of age, gender or culture (Meltzoff1977, Ekman 1993); in social contexts, they influence what we think of a person, and how we behave toward them (Niedenthal et al, 2010); in our own reflection or proprioception, they also influence our moods (Strack, Martin, and Stepper, 1988: Wagenmakers et al, 2016).

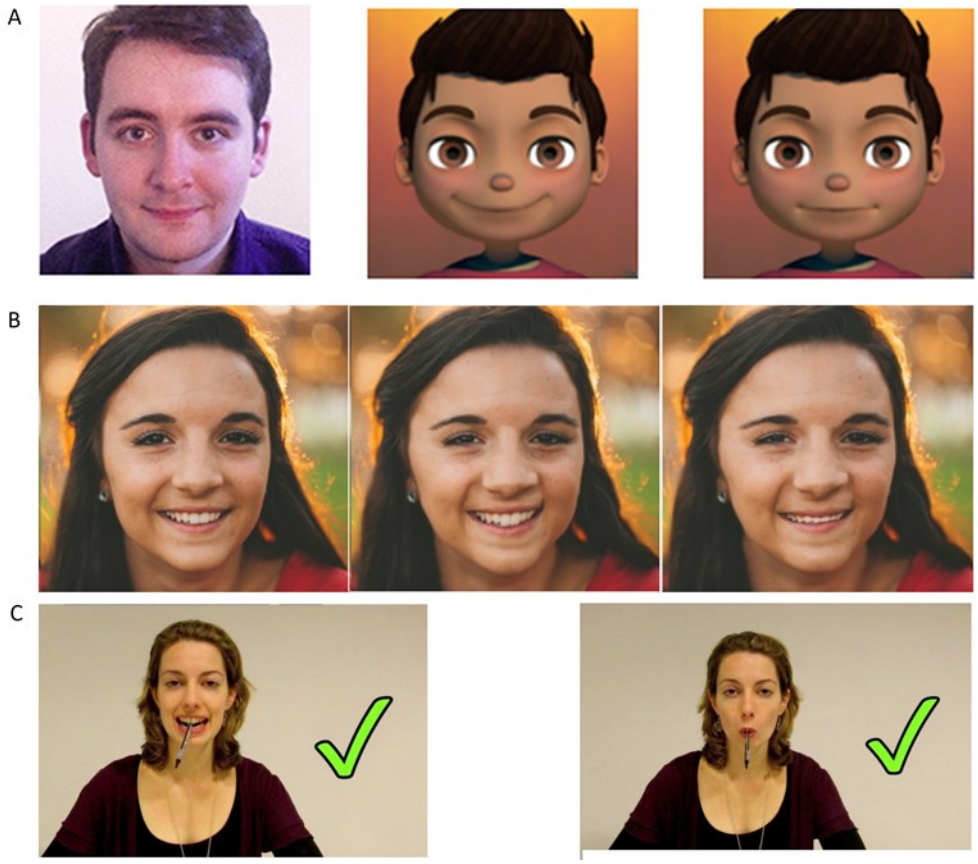

Figure 2. *(A) Common face synthesis systems allow to generate and control smiling expressions in real-time, but only via the mediation of avatars (adapted from Oh et al, 2016; left: original, middle: avatar with enhanced smile, right: suppressed smile). (B) Deep-learning techniques can learn photorealistic transformations of facial expressions on arbitrary photographs, but these techniques do not work in real-time (adapted from Yeh et al, 2016; left: original picture, middle: manipulated picture with enhanced smile, right: suppressed smile). Our proposal of an 'augmented mirror' aims to generate real-time transformations of a user's input, as in A, based on their normal, non-synthetic video stream, as in B. (C) A common experimental manipulation used in emotional feedback experiments : participants are instructed to either hold a pen in their teeth (left; simulating a smile) or between their closed lips (right; simulating a frown). The project will radically renew the field's methodology by using manipulated audiovisual feedback.*

It is therefore not surprising that smiling constitutes a much-researched part of embodied agents designed for human-computer interactions (El Haddad et al., 2016 ; Ochs & Pelachaud, 2017), and that most face synthesis systems for such agents offer some degree of control over their visual and vocal smile behaviour (Yu & Schyns, 2012). However, these techniques mostly allow to synthesize and manipulate the expression of embodied agents, but not to transform the audiovisual expression of real users, e.g. in their reflection in a mirror. In the visual domain, techniques allowing to control the morphological parameters of a synthetic face (e.g. cheek-raising, mouth opening, symmetry, lip press ; Yu & Schyns 2012) can work in real-time, but can only apply to human-human interactions if they are mediated by a virtual avatar (Oh, 2016 – Figure 2A). Conversely, recent deep-learning techniques able to learn expressive transformations from a corpus of paired images (Yeh & Agarwala, 2016 – Figure 2B) allow realistic facial transformations of arbitrary users, but have yet to operate in real-time. Similarly, in the audio domain, hidden-markov models

(El Haddad et al., 2016) and formant resynthesis (Quené, Semin & Foroni, 2012) techniques can reproduce realistic characteristics of speech pronounced while smiling or laughing, but again only in non-real time applications.

Research on realistic real-time emotional transformations has been a recent focus of both partners CentraleSupeler (Soladié et al., 2013, for vision) and IRCAM (Rachman et al., 2017, for speech). Both efforts have recently converged to the same principle of parametric warping based on control points, and the time is now ripe to move beyond the proof of concept and integrate both teams' efforts into a common platform. In more details, the visual part of the algorithm to be developed in project REFLETS will track morphological features of the face, such as the eyes and lip corners, stretch their position using a predefined parametric model, and resynthesize pixel grey levels to correspond to the modified morphology (Figure 3right). Similarly, the audio algorithm developed in the project will track the frequency positions of the vocal formants (bumps and valleys of the vocal spectral envelope), shift their positions and amplitude using a predefined parametric model, and reconstruct the audio signal with the new modified spectral cues.

This approach is original with respect to the state of art in emotional synthesis because it allows to adapt the audiovisual transformations to the morphology of individual observers, and thus create emotional expressions that look and sound like genuine behavior (see Figure 3-left for a recent prototype of the approach). At the same time, it provides an efficient algorithmic framework to reach unprecedented levels of real-time performance, with fast, low-level transformations operating on a frame-per-frame basis, and slower, higher-level feature tracking operating at a longer time scale.

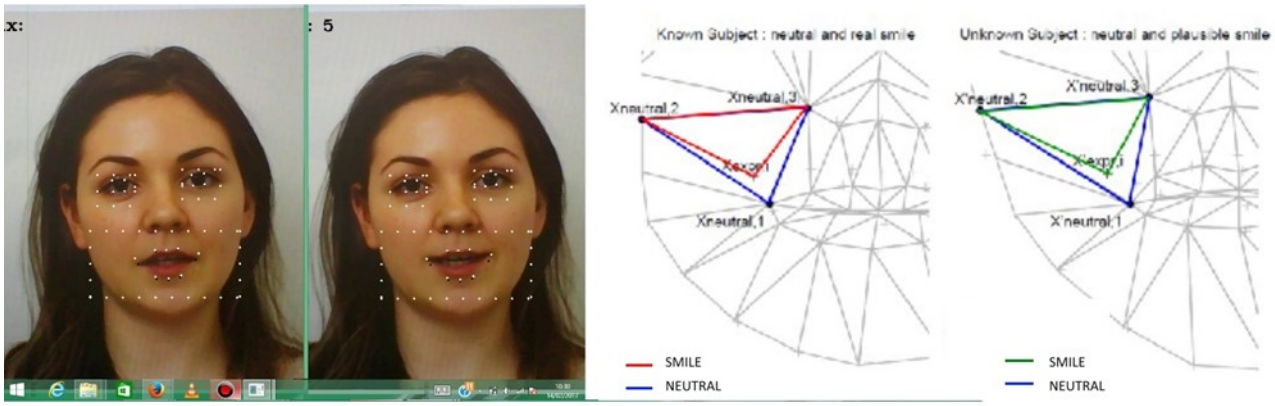

Figure 3*. Realistic smile transformation in real-time video streams. Left : a recent prototype of the approach shows a very promising degree of realism (left : original, right : enhanced smiled). Right : Changes in face morphology due to a given emotional expression (here, smile) can be learned on a dataset of labeled emotional faces (neutral : blue, smiling : red) and parameterized as a linear warping function relative to control points segmented on a subject's neutral face. The warping can then be applied in real-time on control points segmented from an unknown subject, to recreate a similar expression (green).*

I.2.2 Psychological effects of emotional feedback

The relationship between the expression and experience of emotions is a long-standing topic of disagreement in the field of psychology (James, 1894; Schachter-Singer, 1962). Central to this debate have been studies on facial feedback, showing that the surreptitious induction of a smile or a frown leads to congruent changes in the participants' emotional reactions (Strack, Martin, & Stepper, 1988). However, although they support the general notion that emotional expression influence experience, these experiments leave a lot of mechanistic details largely unresolved.

First, in the vast majority of emotional feedback studies, participants were simply instructed or otherwise tricked to express certain emotions in the speech and face. This makes it unclear whether it is the mental processes instigating the action or the afferent feedback from the contracting muscles or the sound of one's voice that created the emotional effect (Buck, 1980; Davis, 2010). The proposal of project REFLETS to study emotional feedback with an augmented mirror is original with respect to this literature, because the

emotional expression will be perceived by the participant without needing any action on his/her part. In a recent experiment, we have used a modified vocal feedback paradigm to show that the mere hearing of one's emotionally transformed voice was sufficient to trigger emotional feedback (Aucouturier et al., 2016), which suggests that this approach is appropriate.

Second, in all studies requiring action from the participant, it is typically very difficult for them to ignore that e.g. they are currently holding a pen in their mouth (Strack, Martin, & Stepper, 1988) and we do not know the extent to which emotional feedback works because, or despite, this awareness of the experimental situation. The goal of project REFLETS to create audiovisual transformations that are undistinguishable from genuine expressions will allow to investigate emotional feedback covertly, with the notable consequence of allowing more efficient within-subject experimental designs and making them usable in repeated therapeutic sessions.

Finally, recent debate on the reproducibility of the pen-in-mouth feedback paradigm (Figure 2C) has suggested emotional feedback effects depend critically on low-level parameters of the experimental situation and participant behaviour. (Wagenmakers et al, 2016). One major problem with previous experiments is that it has been so far impossible to control the intensity of the emotional expression of experimental participants, and more generally link the properties of their expression with a potential feedback effect (should one smile a little or a lot, should one speak while smiling or stay quiet, etc.). By using modified audiovisual reflections, project REFLETS will allow, for the first time, to explore all parameters (lip position, voice parameters, etc.) of what makes a smile self-contagious.

I.2.3 Health and well-being applications of audiovisual smile transformations

While empirical demonstrations of emotional feedback mechanisms have been plentiful in the lab (although non-replications are probably as numerous, see Wagenmakers et al, 2016 and possible reasons why above), there hasn't been a lot of clinical work investigating possible applications of the mechanism in health or well-being contexts.

A series of work have considered clinical cases in which facial expressions were impaired or otherwise controlled and how this modified feedback affected the mood of the patients. Van Swearingen, Cohn, and Bajaj-Luthra (1999) tested a group of people with varying degrees of facial paralysis and found that those who had specific impairment in the ability to smile rated highest on depression scores. With a similar rationale, Finzi and Wasserman (2006) treated 10 people who had symptoms of depression with Botox injections in the forehead – thus artificially paralysing muscles responsible for frowning – and found this improved their depression. While potentially interesting for a variety of psychiatric conditions, the intrusiveness of such setups greatly limit their application. Our proposal with modified audiovisual feedback (seeing yourself smile, instead of paralysing your muscles) constitute a non-intrusive and non-permanent alternative to these treatments, holding potential for larger applicability.

While they were not able to modify a participant's own emotional expression (as we propose to do here), various forms of auditory and visual feedback have also been used clinically, in particular in the context of emotional memory disorders. For instance, Rickard, Wong & Velik (2012) showed that relaxing music played after the exposure to negative stimuli can reduce the consolidation in memory of unwanted material. By introducing a new technology able to automatically modify the tone of voice and facial expression of a patient, our proposal will allow to let participants control the encoding of their emotional memories, directly as they tell them, rather than via an indirect feedback such as background music. Our proposal seems especially relevant to PTSD as the disorder is associated with negative self-referential processing and identity disturbance. In Frewen et al (2011), women with PTSD endorsed more negative and less positive trait-adjectives as self-descriptive, and experienced more negative and less positive affect in response to viewing pictures of themselves. PTSD appear to be one of these disorders where patients clearly need to see and hear themselves more positively (Figure 4).

Figure 4: *Two photographs from the Veteran Vision Project (http://veteranvisionproject.com), an art project by American photographer Devin Mitchell to document the PTSD-ridden inner world of US war veterans, using two photographs composited with a mirror. © COPYRIGHT 2015 DEVIN MITCHELL*

Finally, the project is original with respect to other proposals linked with traumatic stress because it does not target the well-known primary symptoms of PTSD (*the persistent, involuntary re-experiencing of the event* and *excessive avoidance of stimuli associated with the trauma*), but rather the associated symptom of emotional numbing (*loss of interest for previously pleasurable activities* – APA, 2013). Typical PTSD interventions, known as trauma focused cognitive-behavioural therapy (TFCBT) or exposure therapy, involve asking the subject to relive the trauma imaginally, or to expose subjects to cues associated with the traumatic event (for example graded re-exposure to car travel following a road traffic accident). While such therapies are efficient to reduce PTSD symptoms, depression and anxiety (Bisson & Andrew, Cochrane Review, 2005), its focus on trauma is also linked to adverse effects (e.g. increased re-experiencing of the event after re-exposure) and often judged intolerable by patients, leading to problematic drop-out rates. (Pitman et al, 1991). By not focussing on re-encoding extreme trauma-related negative emotions, but rather on reconditioning once-pleasurable, now-bland daily experiences, our approach aims to constitute a non-intrusive, better-accepted alternative to these classical interventions to improve quality of life in PTSD patients. Moreover, as emotional numbing is a risk factor in the general population for emotional disorders such as professional burn out or mild depressive episodes (ex. Mattila et al, 2007), the same approach also holds potential as a well-being device outside of the clinical world.

I.3. Risk management and methodology

The main risk associated with **Objective 1** (audiovisual smile transformations) is that either the visual or the audio algorithmic implementations do not achieve latencies fast enough for sensori-motor feedback (i.e. less than 50ms). This will be addressed either by reducing the realism of the transformations in one or the other modality, or by exploring scheduling strategies (e.g. delaying fast video transformations with respect to slower audio transformations) that benefit from the known temporal tolerance of human observers for audiovisual perceptive fusion (Cavé, Ragot & Fano, 1992).

The main risk associated with **Objective 2** (emotional feedback paradigm) is that the effect size of the intervention is found too small for a clinical application; this will be addressed by changing the primary outcomes of the clinical study, aiming to identify markers of symptom severity rather than a full-remediation of the disorder (see Myers et al, 2016 for a recent example).

The most important risk with **Objective 3** (health and well-being applications) is difficulties to recruit patients in the study (delay, drop-outs, etc.). A typical fall-back solution for clinical studies is to run a similar study on a sub-clinical sample with alexithymia traits – the project already includes this provision, because it will also study well-being applications to a non-clinical population as part of that same objective.

Finally, in the event that, even with such strategies, real-time performance is found impossible at the expected level of realism or self-feedback is found too unreliable for health or well-being applications (a valuable theoretical finding in itself), we will change the 'mirror' use-case to a skype-like dual-video system, for which longer real-time latencies are better-accepted and emotional contagion effect are large and well-

known. Using the same hardware and software device, we will replace one-person interaction with the mirror to a two-person interaction with a therapist. Even in this situation, the ambitions of the project to provide new, innovative health technology with an important potential for valorisation as a well-being device will be fully preserved.

II. Project organisation and means implemented

II.1. Scientific coordinator

Computer scientist **Catherine Soladié** will coordinate this proposal. Before becoming Associate Professor at CentraleSupélec in the FAST team, Catherine Soladié was project manager for 4 years at CapGemini. The project will directly benefit from her organizational talents as a project coordinator, as well as her skills in emotions analysis and synthesis as the coordinator for real-time face synthesis research (SP1). Catherine Soladié was also part of one of the two Immemo project teams (ANR CONTINT project coordinated by FAST) who won the first two places in the international challenge AVEC 2012 in emotion analysis. This same Immemo project won the first place in the International Challenge FERA 2011 on micro expressions detection.

II.2. Consortium

IETR

Project management, technological developments for visual (facial) emotional transformations as well as valorisation and market research will be handled in project REFLETS by the FAST (face analysis and synthesis) team of UMR6164 IETR (CNRS/CentraleSupelec), located in Rennes (France).

FAST has over 15 years experience working with video facial analysis, contributing new algorithms to optimize analysis speed for deformable models of facial morphology, in particular in the context of emotional expressions. FAST was the coordinator of ANR Immemo (ANR CONTINT 2009), who won first place in two international challenges in domains closely related to project REFLETS: first place for detecting facial micro-expressions (FERA 2011 ; see Sénéchal et al., 2012) and first and second place for emotion recognition in audiovisual signals (AVEC 2012 ; see Soladié et al., 2012). FAST is also especially experienced in project valorisation in terms of startup companies. FAST research activities in the past have lead to the offshoot of two successful startup companies : 3D Sound Labs & Dynamixyz, the latter also involved in the present consortium (see below).

IRCAM

Technological developments for audio (speech) emotional transformations, as well as cognitive psychology experiments in emotional feedback will be handled in project REFLETS by the Perception and Sound Design team of UMR9912 STMS (CNRS/IRCAM/UPMC), located in IRCAM (Institut de Coordination Acoustique/Musique) in Paris (France). IRCAM was founded by composer Pierre Boulez in 1977; it is now the world's largest research institute in computer music, as well as the country's only laboratory fully devoted to the science and technologies of music and sound. Fully equipped with psychoacoustics and neuroscience experimentation booths, IRCAM's Perception and Sound Design team is the only research unit in the institute devoted to cognition and experimental science of sound and music. Taking roots in the seminal studies of music timbre perception by David Wessel and Steve McAdams, work in the Perception team now encompasses topics as varied as sonic environmental quality (for which we won the Ministry of Environment's Décibel d'or 2014 award), sound design (we designed sound for the new Renault electric cars) and music neuroscience. IRCAM brings to the project high expertise in emotional speech transformation technologies, as well as in the cognitive psychology of emotional feedback (see e.g. https://www.youtube.com/watch?v=jdneqc-i18g). Jean-Julien Aucouturier (neuroscientist, chargé de recherche CNRS, PI ERC StG CREAM) and Marco Liuni (computer-scientist, chargé de recherche IRCAM) will coordinate IRCAM participation in the project.

Cognac G

Clinical applications of the augmented mirror to PTSD patients will be handled in project REFLETS by UMR 8257 Cognac-G (CNRS/Paris-Descartes/Service de Santé des Armées), located at Percy Military Hospital in Clamart (France). Cognac-G works with ecological pervasive environments (e.g. flats, cockpits), smart sensors and applied mathematical tools, in order to measure normal and abnormal human behavior. The unit will bring its clinical expertise and infrastructure to the project, notably its capacity to recruit civil and militarian patients for longitudinal studies and access to the 'smart flat' platform at Ecole du Val de Grace (Paris) and Percy military Hospital, where the studies will be conducted. Stéphane Buffat (Médecin en chef IRBA) will coordinate the Cognac-G participation in the project, with a team also consisting of clinical neurologists Damien Ricard and Flavie Bompaire.

HumanEvo

Well-being applications of the augmented mirror for the non-clinical, general population will be handled in the project by Human Evo, a coaching private practice specialized in behavioral and cognitive therapies (BCTs), located in Neuilly (France). Human Evo works on human factors in business and institutional environments, and collaborates on a regular basis with the French National Police Academy, a context in which PTSD and emotional regulation issues are common concerns.

Dynamixyz

Camera calibration in the augmented mirror hardware prototype will be handled in the project by Dynamixyz, a software company located in Rennes (France). Dynamixyz is the market leader in performance capture technology for the entertainment industry (film and AAA video games), and offers solutions to transfer an actor's emotional expressions onto a 3D character and animate 3D talking heads with realistic emotional expressions. The company will be in charge of detecting a hundred points in 3D on the user's face from multiple cameras behind the mirror, in order to support the visual smile transformations developed in the project.

Chanel

Hardware development of the project's augmented mirror platform will be handled by the BWB Team (Biology and Women's Beauty) of the Chanel company. Chanel's BWB team works on the perception of aging, attraction and health returned by the face related to both genetic and psychological factors. Specifically, the "Face Perception" unit is specialized in cognitive psychology and studies self-perception, self-esteem and perception of health. Chanel's BWB team will develop the mirror hardware used in the project to analyze and re-synthesize the face and speech of the observer.

II.3. Means of achieving the objectives

Work planned under project REFLETS will be organized in 5 sub-projects: the project consortium will collaborate to create the mirror hardware (SP2), develop audiovisual smile transformation algorithms that will be embarked in the mirror(SP3), conduct human experiments, both in the lab, and in the context of health and well-being applications (SP4) and work towards a societal valorisation of the work, in the form of a startup company (SP5).

SP1 : Project Management							Coordinateur IETR/CS
Partenaires	IETR	IRCAM	Dynamixyz	HumanEvo	CognacG	Chanel	p.mois
p.mois	4	1	1	1	1		

Objectives, work program and contributions

The objective of the SP1 is to ensure proper management of the project, to promote collaboration dynamics between the partners and to produce 'consortium agreement' documents in a quick and efficient manner. Project coordination will be ensured by periodic advancement meetings (steering committees every six months), with the addition of technical meetings that will be essentially bipartite. The drafting of the

'consortium agreement' documents will start at the beginning of the project with the objective of being finalized at T0 + 12. Special attention will be devoted to this agreement, which will be used to fix the rules of collaboration during the project but should also plan for future valorization efforts, taking into account the interest of the project's academic and industrial partners in the commercialization of the result of the project.

Deliverables
L1.1.1 à L1.1.7 (IETR) : Meeting reports.
L1.2 (IETR, T0+12) : Consortium agreement.

SP2 : Hardware							Coordinateur Dynamixyz
Partenaires	IETR	IRCAM	Dynamixyz	HumanEvo	CognacG	Chanel	p.mois
p.mois	6	0	42	0	1		

Objectives, work program and contributions

Industrial partner Chanel will be responsible for the creation of a tinted mirror which will give the possibility to illuminate the face of the person so that it can be seen through a computer screen. Work by Chanel will take into consideration the point of views of Human Evo, IRCAM and IRBA concerning the ergonomics of the device, as well as that of Dynamixyz who will handle the cameras.

A first mirror is already working (see figure 1) and will be used by the partners in SP4 from the beginning of the project.

Dynamixyz will be responsible for cameras calibration in real time and on the fly. The result of the calibration will be illustrated by the production of a 3D representation of the face of the person facing the mirror. High-quality calibration on the fly is needed to provide reliable control points for subsequent transformations of facial expressions.

Deliverables
L2.1 (Chanel): Mirror delivery with only one camera: M1.
L2.2 (Chanel): Mirror delivery with four cameras: M2 with an SDK to retrieve the images captured by the cameras and to manipulate the LEDs.
L2.3 (Dynamixyz): On the fly calibration SDK, production of a 3D representation of the person's face.

SP3 : Emotional Software							Coordinateur IETR
Partenaires	IETR	IRCAM	Dynamixyz	HumanEvo	CognacG	Chanel	p.mois
p.mois	42	15.3	8	0	1		

Objectives, work program and contributions

IRCAM has already developed a software proof-of-concept able to modify the positive component of perceived emotion in a person's voice message (Arias, Belin & Aucouturier, 2017). While this prototype is compatible with a frame-level real-time implementation, the algorithm currently doesn't operate in real-time. IRCAM will work in SP3 to improve the realism of the algorithm and adapt it for real-time performance.

IETR already has the capacity to modify facial expressions in real time video streams. As part of the project, the underlying algorithms will be improved by taking into account the 3D shape of the face produced by Dynamixyz as part of SP2.

IRCAM and IETR will also collaborate to implement both algorithms into a common multimodal pipeline, increasing the efficiency of the implementation and benefiting from cross-modal information to improve the functionality of each block. In particular, both algorithms will be extended to adapt to the actual

expressivity of the participant. At present, if a participant speaks with a joyful tone of voice, the voice transformation algorithm will automatically produce an even-more joyful voice without consideration for the participant's initial emotional state. In the case of voice, it will therefore be necessary to model the neutral expression of the person in order to manipulate it *relatively* in the direction of more or less positivity. Similarly, it will be necessary to model the neutral facial expression of the person in order to synthesize a more smiling face. To capitalize on the audiovisual nature of the project, neutral state detection will be done jointly in the audio and video domain.

Note: IRCAM and IETR are already working together on the production of a real-time software making it possible to produce a smiling 2D face and a more positive voice (see prototype in Figure 3-left). We are confident in the fact that a minimalistic real-time implementation of the system will be available only a few months after the beginning of the project. This will allow the experiments of SP4 to begin in time, without depending on more complicated developments, which can be incorporated at a later stage.

Deliverables

L3.1 (IRCAM): Real Time software in order to change in a positive manner the expressiveness of a person voice.

L3.2 (IETR): Real Time software in order to change in a positive manner the expressiveness of a person face, taking into account the 3D modeling of the analyzed face..

L3.3 (IETR): Real Time software realizing the preceding process, adapted to the present audio and video expressiveness of the person.

SP4 : Experiments							Coordinateur IRCAM
Partenaires	IETR	IRCAM	Dynamixyz	HumanEvo	CognacG	Chanel	p.mois
p.mois	2	40	1	10	22		

Objectives, work program and contributions

SP4a. Cognitive psychology (IRCAM)

Experiments in SP4a will use the audiovisual smile transformations developed as part of SP2 and SP3, to conduct a series of laboratory psychological experiments and establish the methodological conditions of a strong emotional feedback effect amenable to the clinical application of objective 3 (below). Project partner IRCAM will coordinate the studies.

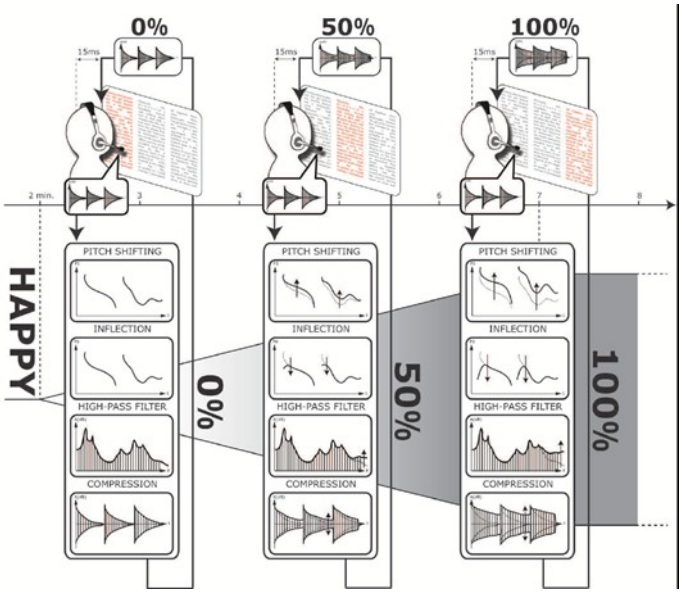

Figure 5: *Audiovisual feedback paradigm (here, illustrated on a happy voice) in which participants are asked to read a text while their feedback is continuously manipulated to sound or look more emotional. Figure adapted from Aucouturier et al, 2016.*

We will adapt the experimental paradigm of Aucouturier et al., 2016, in which participants were tasked to read a text out loud while their voice was algorithmically modified to sound more happy or sad. In this proof of concept, the emotional state of the participant was measured with subjective adjective scales before and after reading, while their physiological arousal was measured continuously while reading using skin conductance level.

Building on this proof of concept, SP4a will study the following methodological and theoretical elements of the emotional feedback mechanism:

1. Audiovisual feedback: the experimental paradigm will be extended to the audiovisual case, in which participants can both see their manipulated (smiling) reflection and hear themselves with a more positive tone. Using both congruent and non-congruent feedback (ex. Smiling face with a neutral voice, neutral face with a smiling voice), we will establish which expressive channel is most important for emotional feedback and how these two modalities interact.
2. Experimental task: in Aucouturier et al. 2016), participants were instructed to read a text (a novel by Japanese writer Haruki Murakami), and this was found sufficient to create an emotional feedback effect. However, participants did not attribute modified emotional meaning to the text itself (they only reported that their own subjective state had changed). Using audiovisual feedback, we will investigate a number of other situations more suited to the health and well-being applications of SP4b and SP4c, incl. reading a text that they wrote themselves, recollecting a personal memory, verbally describing the content of an emotional picture and freely interacting with an experimenter.
3. Effect of duration: the emotional feedback effect found in Aucouturier et al. 2016 was measured after 10 minutes of exposure. In order to determine the best user scenario for the applications of SP4b and c, experiments in SP4a will investigate how the size of the effect varies with duration and repetition of the task (e.g. shorter sentences, separated with neutral blocks).
4. Effect of awareness: In our pilot study, participants did not consciously detect the vocal manipulation. However, it is unclear whether the effect depends on this lack of awareness – can it only occur below the threshold for detection (i.e. with small, undetectable smile transformations and a cover story for the need of audiovisual feedback) or can it be improved with stronger manipulations even if the participant is aware of the manipulated feedback.

Experiments in SP4a will be conducted by the Perception and Sound Design team of UMR9912 STMS, using an experimental platform made available via University partnership in the context of COMUE Sorbonne Universités (the INSEAD/Sorbonne Behavioural Center http://centres.insead.edu/sorbonne-behavioural-lab/fr/etude/index.cfm). The INSEAD Centre offers verified management of participant recruitment, consent, data anonymity and payment, as well as complete agreement with the European ethical regulations for experimental research. In particular, all experiments in SP4a will be ethically screened and authorized by the INSEAD Institutional Review Board (IRB) before participants are included.

SP4b – Clinical application to PTSD patients (Cognac-G)

The study in SP4b aims to assess the ability to test clinically the positive self reflection as a Complementary and alternative medicine (CAM) for post-traumatic stress disorder symptoms. The study will follow a patient-control paradigm, in which participants will be either patients with diagnosed PTSD, or healthy people matched in age and gender, such as described in the inclusion and exclusion detail below. Partner Cognac-G, a clinical research unit co-operated by the Service de Santé des Armées, will coordinate the study.

Inclusion and exclusion criteria for patients:

Adults patients diagnosed with PTSD caused by mild brain traumatic injury and/or adult participants who will have been administered a measure assessing PTSD symptoms. Although generalisation among patients with PTSD caused by other causes can be discussed, the choice we make is both due to the

availability of a homogeneous cohort. Participant with visual (uncorrected) or auditory deficit will be excluded.

Measures of PTSD symptom severity:

Upon inclusion, we will measure the severity of PTSD symptoms in patients and controls using the Jellinek-PTSD screening questionnaire *(*JPSQ), a short self-report questionnaire and serves as a first screening instrument to identify participants which might suffer from PTSD (van Dam et al., 2013). The instrument consists of four questions that can be answered with either yes or no. The score is the total sum of positive answers (range 0–4). The JPSQ has shown to have high sensitivity (0.87) and specificity (0.75) (van Dam et al., 2013). In addition, the French version of the BDI-II questionnaire (Freeston et al., 1994) will be used to assess the presence of depressive symptoms in the past 2 weeks. The BDI-II-NL consists of 21 statements that can be answered on a four point scale with higher scores representing more depression. Finally, the STAI-DY questionnaire will be used to assess trait anxiety prior to the experimental manipulation. The STAI-DY consists of 20 items, which can be scored on a four point scale, with higher total scores indicating higher levels of trait anxiety.

Clinical outcome assessment:

Usual PTSD symptoms such as intrusion, avoidance, negative alterations in cognition and mood, and alterations in arousal and reactivity will be assessed by the project's psychiatrist, and the patients will be rated by means of the Diagnostic and Statistical Manual of Mental Disorders (DSM V). In addition, we will perfom a questionnaire battery similar to that used by Dibbets & Schulte-Ostermann, 2015. Mood changes following the intervention will be assessed by means of the French version of the Profile of Mood States (P.O.M.S.) ratings questionnaire. It consists in sixty-five items. Each item can be scored on a Likert-scale, ranging from 0 (not at all) to 4 (extreme). Higher scores indicate a higher negative mood. Finally, a dedicated acceptability questionnaire will be used. It will contribute to the global clinical evaluation, and will also be used in the general population assessment of SP4c.

Location:

All experiments in SP4b will be conducted by Cognac G in the Smartflat platform, situated in the neurological service of Percy Hopital d'Instruction des Armées. The platform is a hospital room refurbished as a small apartment but retaining all the necessary support of the original hospital room. In addition, it is equipped with an array of sensors, and a contiguous control room with the hardware, software and datastorage support. This location will enable a range of complementary measures of participant behaviour based on sensor data (e.g. video cameras, microphones, ambient and portable eye-traking devices), which can be used as secondary measures of the effect of the intervention. All clinical data will be recorded and stored at Percy Hospital in compliance with ethical and data safety requirements.

SP4c - Emotional well-being in the general population (HumanEvo)

In addition to the formal clinical study of SP4c, the potential of the project's augmented mirror as a well-being device in the general, non-clinical population will be assessed using a user study, using the framework of psychosocial interventions known as Cognitive Behavioral Therapies (CBT). Partner HumanEvo, with more than 15 years experience in human coaching and CBTs, will coordinate these tests.

User tests will follow the format of a series of brief BCT sessions, each consisting of three steps :

- initial briefing between the participant and the experimenter/therapist (session objectives, assessment of the initial emotional state)
- participant uses the device (with or without the experimenter's intervention) to record a video log of themselves describing a midly pleasing, neutral or unpleasing (past or anticipated) experience.
- Final debriefing (what the participant considered effective in the session, assessment of final mood state)

Experience described in front of the mirror	***Expected effect of the smiling feedback***
Past event, pleasing or neutral (e.g. how I went to the movies last week-end)	Feeling like re-experiencing the event, and thus to seek the experience again
Anticipated future event, pleasing or neutral	Increased motivation to experience the

(e.g. going to the movies next week)	event
Past event, midly unpleasing (e.g. embarrassing myself in public at work yesterday)	Increase perspective taking, decreased emotionality in re-telling the event.
Anticipated event, unpleasing (e.g. a public presentation I have to give next week)	Reduced anxiety towards the event, improved coping and decreased emotional response during the event

Before the initial session, each user will fill a questionnaire allowing them to report possible events of interest, along with a self-evaluation of emotional intensity. A series of sessions, designed to focus on events of incremental emotional significance, will be determined on an individual basis by the experimenter.

Deliverables

L4.1 (IRCAM) – Report on cognitive experiments, in the form of open access journal publications or open-access preprints

L4.2 (IRCAM) – Complete, anonymous experimental data in the form of internet data repository (e.g. figshare.com) and code for statistical analysis (Python notebooks)

L4.3 (Cognac-G) – Report on clinical study, in the form of open access journal publications or open-access preprints

L4.4 (HumanEvo) – Complete minutes of the initial and final briefings of the user tests

SP5 : Valorisation							Coordinateur IETR/CS
Partenaires	IETR	IRCAM	Dynamixyz	HumanEvo	CognacG	Chanel	p.mois
p.mois	22	4	2	1	1	3	

Objectives, work program and contributions

Partners in project REFLETS have the ambition to create a product/device actually used by people with emotional regulation issues, such as those related to alexithymia, burnout or mild depression, both in the clinical and the non-clinical general population. The format of valorisation currently favored by the partners is that of a startup company, building on the proof-of-concept done in SP4. Partner IETR, with experience in already establishing three successful startup companies (Dynamixyz, 3D Sound Labs and ImmersiveTherapy) based on research results in audio and video signal processing, will coordinate this SP.

The first objective of SP5 is to define what market is addressable for what product. Three markets are identified so far:

1. The luxury retirement home market, with which one of the partners is already in contact. This would be a B2B business
2. The general public market. While it would be excluded to promote the mirror hardware developed within this project, a version on tablet and smartphone (Androïd and IoS) would be relevant. This poses the problem of real-time processing, but such implementations are part of the road map of the two project partners who provide the technologies of voice and images analysis.
3. The corporate market for communication software frameworks. A plugin for Skype or any other means of internal communication (ex. videoconferencing) could interest major groups, in particular those concerned with the prevention of risks associated to human resources (e.g. emotional fatigue in call-centers) or improved customer experience (e.g. video customer feedback for customers with hearing disabilities).

The second objective of SP5 will be to identify the competitors in each of the three markets mentioned above with an analysis of their specific business model.

The last objective will be to create a Business Plan in order to compete for French ILAB financing (*Concours national d'aide à la création d'entreprise innovante- BPI*).

Deliverables

L5.1 (IETR) – Market analysis report
L5.2 (IETR) – Competitor analysis
L5.3 (IETR) – Business Plan

As can be seen in the Gantt diagram in figure 6, the project will have a duration of 42 months. Testing in cognitive psychology (SP4a) will start early, using an operational mirror (M1) with a single camera on which IRCAM and IETR have already started working. A second version of the mirror (M2) will be used for clinical testing. The work on valorisation will lead to the proposal of a business plan at the end of the project.

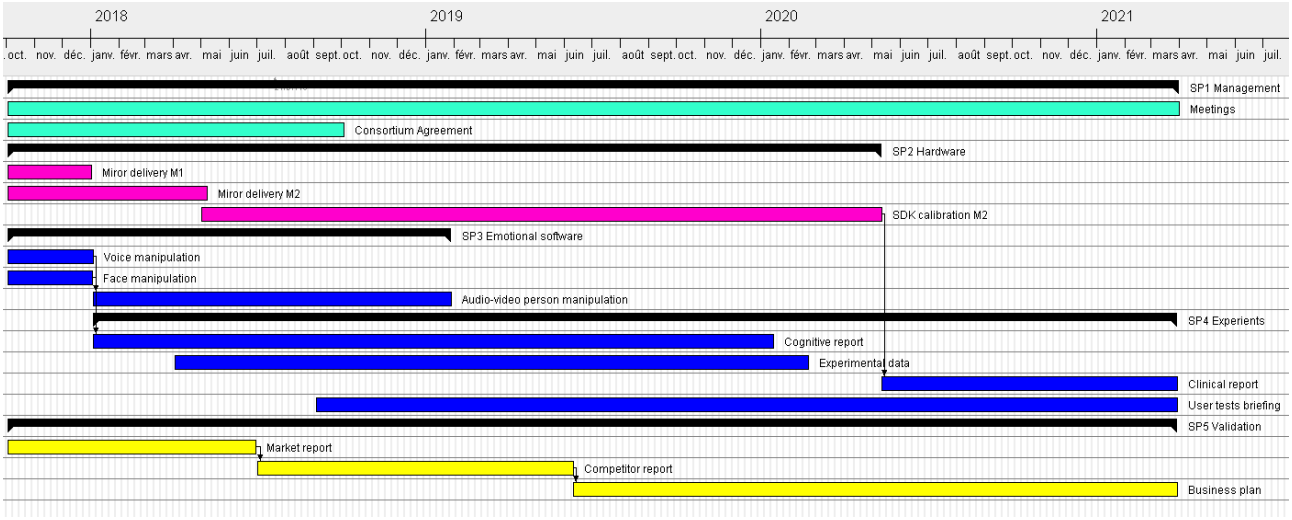

Figure 6*. GANTT diagram for project REFLETS.*

Financial aspects

The grant request (about 660K€ for a period of 42 months) will allow IRCAM (170K€) and IETR (160K€) each to hire a PhD student in the project and to participate to international conferences, plus a seven months post-doc for IETR. Dynamixyz (138K€) will dedicate the work of two R&D engineer to realise the work on real time calibration. HumanEvo (104K€) will spend time to monitor people during the test phase. Chanel (37K€) will adapt a mirror design in order to supply it to the project and finally Cognac G (51K€) will make available the SmartFlat with the associated service.

Previous or ongoing projects and funding received in connection with this proposal:

Title of the call for proposals, source of funding	Project title	Name of coordinator	Starting date/End date	Grant amount	Name	Person. Month
					Of the person involved in this proposal	
ERC Starting Grant	CREAM	JJ Aucouturier	Oct 2014-Oct 2019	1,5 M€	JJ Aucouturier	15mm

III. Impact and benefits of the project

III.1 Impacts

The **scientific impact** of the project is two-fold: first, by producing the first multimodal model of audiovisual deformation, it will provide an algorithmic platform to investigate what time constraints are involved in synchronous vocal and facial feedback – a question important to the field of multimodal interfaces and virtual/augmented reality, notably in time- and performance-critical applications such as those investigated at IRBA. Second, by also opening up the mirror's software tool to the cognitive psychological scientific community, we will provide a reproducible experimental platform to foster collective progress on the much-debated theory of emotional feedback.

The **societal impact** of the project stems, first, from its focus on PTSD, which is a growing concern in modern western societies with e.g. the post-deployment health of army personal as well as civil survivors of unpredictable terror attacks (see e.g. Dantchev et al, 2016 about November 13, 2015 in Paris). Second, the health technology developed in the project will find applications beyond PTSD, for preventive (e.g., a daily, emotional reinforcement of positive experiences) or remediative mental health paradigms (e.g., treatment of anhedonia in major depressive disorders), and as a help module for real-time communication in couple/family therapy.

III.2 Issues

We believe that the requested financing instrument is well-adapted to this project:

- the two academic laboratories (IETR and IRCAM) use very complementary techniques (audio versus video) and therefore can but enrich themselves with this collaboration;
- the project's disruptive potential stems from its organizing a rare confrontation between medical doctors with researchers in signal processing. These two communities can not carry out this project alone, which is the very essence of a collaborative project.

We submitted the project as part of Axis 12 of Challenge 4 (health technologies) for several reasons. While its partial focus on well-being applications in the general population could have made it a part of other Challenges (e.g. axis 8 of challenge 8 concerning public health), we deemed that the project's most specific characteristics was to strongly imply patients and to lead to an industrial transfer given its high potential for technological and usage innovation.

III.3 Valorisation

Academic valorisation

Academics will publish their results in international journals and conferences in order to confirm their position in the field of emotions analysis and synthesis. These publications will address several domains: signal processing (*IEEE transactions*) as well as specialized journals in cognitive psychology (*Emotion, Psychological Science*) or in the psychiatric field (*Stress, Journal of Clinical Psychiatry)*. More globally, strong results about emotional feedback, especially when using innovative experimental paradigms as we do here, will plausibly be considered for publication in high-impact generic journals (see e.g. Aucouturier et al, PNAS 2016).

General public dissemination

The results of the project will be presented at local public events (Laval Virtual), regionally (Fête des Sciences in Rennes Space for SMEs), nationally (Science Festival, on the stand of CentraleSupélec) and Internationally (SIGGRAPH Emerging technologies).

Industrial valorisation

The FAST team of IETR is currently co-founder of three startups Dynamixyz, 3D Sound Labs and ImmersiveTherapy. Dynamixyz (a SME of ten people created in 2010) offers solutions in the field of cinema and AAA video games. This startup was created thanks to the collaborative project IMMEMO (ANR 2009) coordinated by FAST. 3D Sound Labs (a SME of fifteen people created in 2014) work in the domain of 3D sound for headphones and takes advantages of a technology transfer from FAST in the domain of deformable models. Last, Immersive Therapy created this year comes from collaborative work between CognacG and IETR.

BUilding on this successful valorisation experience, team FAST plans to create a startup in collaboration with other partners of REFLETS based on the results of the project if they are positive

Education valorisation

Entrepreneurship is an integral part of the curriculum at CentraleSupélec, an educational institution in which members of the IETR FAST team work. ¨Project members Catherine Soladié and Renaud Séguier both teach a class at CentraleSupélec dedicated specifically to startup creation and societal impact. . About fifty students over three years will be called upon to work on the SP5.

Accord de consortium

We will use the reference model of the consortium agreement on the ANR website: unicANR. We will take great care in defining common results. In particular, the co-owners will decide whether or not to file patents. We will nevertheless favor filing single-partner patents for the sake of simplification. We will define in this agreement all the conditions favoring the technolocal transfer in the form of licensing or redemption of IP so that the researchers, in particular academics, will be remunerated in an equitable way according to their contribution.
Finally, the objective of this project is to establish lasting links between the partners: the post-project can be the object of collaborations between the startup and the project's academic members in the form of Cifre contracts.

Note: as part of the valorisation effort, a patent concerning an algorithmic framework to simulate the acoustics of smiled speech has already been filed by project partner IRCAM (in co-ownership with CNRS) in Feb. 2017 (France N° 1751163).