

# Synthèse personnalisée de l’expression de joie

Sarra ZAIED, Catherine SOLADIÉ, Pierre-Yves RICHARD

CentraleSupélec \*, IETR, Équipe FAST

Campus de Rennes, Avenue de la Boulaie, 35510 Cesson-Sévigné, France

Sarra.Zaied@centralesupelec.fr, Catherine.Soladie@centralesupelec.fr

Pierre-Yves.Richard@centralesupelec.fr

**Résumé** – La théorie de la rétroaction émotionnelle périphérique est le fait que nos expériences émotionnelles soient sous l’influence rétroactive de nos propres expressions. Ce sujet est 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. Par ce travail, nous proposons une nouvelle technologie de transformation visuelle permettant de canaliser le potentiel du mécanisme psychologique de rétroaction émotionnelle faciale pour une application clinique dans le domaine du stress post-traumatique (PTSD). Notre système est capable de transformer de manière photo-réaliste et en temps réel un visage détecté en un visage plus souriant. La particularité est que nous synthétisons une expression de joie spécifique à la personne, tout en gardant l’identité de l’émotion. Les résultats qualitatifs et quantitatifs démontrent que notre méthode spécifique à la personne peut générer des expressions de joie personnalisées plus proches de la vérité terrain que celles synthétisées avec deux approches récentes de l’état de l’art.

**Abstract** – 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 in the 19th century. 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. In this paper, we propose a new health technology able to channel the psychological mechanism of facial emotional feedback for clinical application of post-traumatic stress disorders (PTSD). Our system transforms in real-time a detected face into a personalized joyful one. The particularity is that we synthesize a person-specific joy expression, while keeping the identity of the emotion. The qualitative and quantitative results demonstrate that our person-specific method can generate personalized joy expressions closer to the ground truth than two generic state-of-the-art approaches.

## 1 Introduction

Le sourire est un outil de communication non verbale qui provoque une réduction du stress et une brève sensation de détente et de bonheur [1]. Des chercheurs en psychologie ont montré que le mouvement des muscles faciaux influe sur les expériences émotionnelles [2]. En voyant quelqu’un sourire, nos neurones miroirs s’activent et créent dans notre esprit le sentiment associé au sourire [3]. Pour agir positivement sur l’émotion d’une personne, nous proposons un prototype en temps réel dans lequel les observateurs se voient de manière progressivement plus positive. À leur insu, leurs visages réfléchis sont transformés de façon algorithmique pour apparaître plus souriants sur un ”faux” miroir commandé par ordinateur. Notre méthode est actuellement en phase de test par l’hôpital militaire Percy (France) afin de remédier aux troubles émotionnels chez les personnes souffrant de syndrome de stress post-traumatique (PTSD). Nous nous attendons à ce que notre système agisse sur leurs capacité à percevoir et réguler leurs propres émotions.

Certains chercheurs ont travaillé à la synthèse d’expressions faciales dans différents contextes d’application tels que l’amélioration de la créativité et de la communication dans le cadre du

travail collaboratif [4]. Des outils de synthèse d’expression faciale ont été utilisés pour manipuler et évoquer des émotions [5, 6] en se basant sur des méthodes géométriques. La principale limite de ces recherches est qu’elles utilisent toujours le même modèle avec les mêmes paramètres pour générer un sourire générique. Les déformations appliquées sur les visages sont identiques pour tous les individus. Cependant, chaque personne a sa propre façon de générer des expressions.

Récemment, les modèles génératifs ont obtenu des résultats impressionnants dans les applications de synthèse d’images [7]. Wang et al.[8] propose un UC-GAN pour synthétiser les expressions faciales. Le réseau proposé peut générer des expressions naturelles et réalistes mais pas personnalisées. En effet, ces modèles génératifs apprennent différentes manières de sourire à partir de toutes les personnes de la base d’apprentissage. Malheureusement, sans aucune connaissance préalable d’un nouveau sujet de la base de test, le GAN ne peut pas deviner la manière dont le nouveau sujet réalise son sourire. Ainsi, les sourires générés sont réalistes et différents d’une personne à une autre mais ne sont pas les expressions réelles de chaque personne.

La contribution principale de cet article est de proposer une approche géométrique qui permet de générer l’expression de

\*Travaux financés par le projet ANR REFLETS.

joie réelle de la personne. L'originalité de notre méthode est que nous apprenons un modèle paramétrique spécifique pour chaque personne grâce à une connaissance préalable de la façon dont la personne réalise son sourire. Le modèle est appris en utilisant une image neutre et une image souriante du sujet. Ce modèle est ensuite utilisé pour synthétiser une expression plus souriante spécifique à la personne au cours du temps sur une image quelconque de ce même sujet. En nous basant sur une détection en temps réel des points caractéristiques [9] et une méthode de warping [10], nous pouvons générer des sourires photoréalistes et personnalisés tout en préservant l'identité de l'utilisateur et l'identité de son émotion. La deuxième contribution concerne les propriétés du modèle, qui nous permet de synthétiser des sourires avec des intensités différentes. Nous comparons nos résultats avec deux approches récentes : une méthode géométrique [6] et une méthode générative [8].

## 2 Notre méthode

Notre méthode fait partie des méthodes géométriques. Elle consiste à apprendre un modèle spécifique à la personne en utilisant une image neutre et une image souriante (cf 2.1) comme le montre la Fig. 1. Elle nous permet de synthétiser des sourires avec des intensités différentes (cf 2.2) en se basant sur une méthode de déformation 2D "MLS" [10].

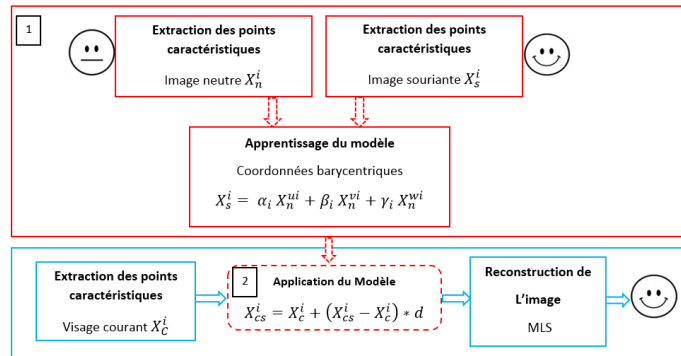


FIGURE 1 – Le système est composé de 2 parties. En bleu : la méthode géométrique [6] basée sur l'extraction de points caractéristiques du visage et la méthode MLS [10] qui permet de déformer le visage détecté. En rouge, nos contributions : 1/ L'apprentissage d'un modèle spécifique à la personne à partir de son visage neutre et son visage souriant. 2/ L'utilisation d'un coefficient qui permet de varier l'intensité de l'expression de joie synthétisée.

### 2.1 Apprentissage du modèle spécifique à la personne

La première étape de notre méthode consiste à apprendre un modèle paramétrique spécifique à la personne en utilisant une image neutre  $X_n$  et une image souriante  $X_s$  de cette personne. Pour la détection des points caractéristiques, nous uti-

lisons GenFaceTracker de Dynamixyz [9] qui détermine avec précision les coordonnées de 84 points caractéristiques dans chacune des images comme l'illustre la Fig 2.

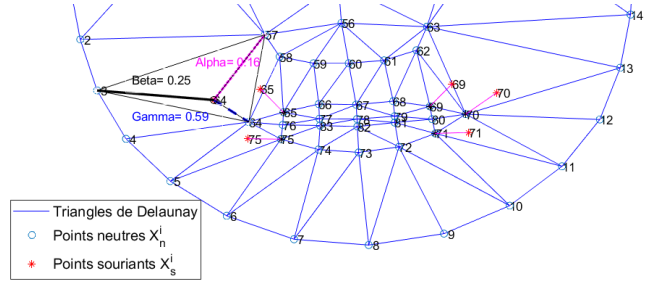


FIGURE 2 – Les positions des points souriants  $X_s^i$  par rapport aux triangles du visage neutre  $X_n^i$ . En souriant, chaque point de contrôle du sourire est situé dans un triangle du visage neutre ( $X_s^{64}$  est dans le triangle des sommets  $X_n^{57}, X_n^3, X_n^{64}$ ).

Un sourire est exprimé avec la montée des coins de la bouche et des joues, ainsi que la levée des paupières inférieures [11]. Nous avons sélectionné les 10 points caractéristiques correspondant aux coins de la bouche et aux points inférieurs des yeux pour apprendre et appliquer les déformations. Nous alignons les deux visages  $X_n$  et  $X_s$  à partir des points du contour du nez et des sourcils car la position de ces points ne varie pas beaucoup entre les 2 images. Nous effectuons une triangulation de Delaunay sur le visage neutre. Chaque point de contrôle du sourire  $X_s^i$  est situé à l'intérieur d'un triangle du visage neutre ( $X_n^{u_i}, X_n^{v_i}, X_n^{w_i}$ ) comme le montre la Fig. 2 pour le point 64 du visage souriant. Nous calculons les coordonnées barycentriques  $(\alpha_i, \beta_i, \gamma_i)$  pour chacun des 10 points du visage souriant  $X_s^i$ . Ces coordonnées sont les paramètres de notre modèle spécifique à la personne. Le modèle personnalisé est donc composé de 10 vecteurs à 6 composantes. Ces composantes sont les 3 index de sommets de triangles ( $u_i, v_i, w_i$ ) et les 3 coordonnées barycentriques  $(\alpha_i, \beta_i, \gamma_i)$  associées à chacun des 3 sommets. Le calcul de  $X_s^i$  est formulée comme suit :

$$X_s^i = \alpha_i X_n^{u_i} + \beta_i X_n^{v_i} + \gamma_i X_n^{w_i} \quad (1)$$

### 2.2 Génération de l'expression de joie

Une fois que le modèle spécifique est appris, nous pouvons modifier en temps réel chaque nouvelle image  $X_c$  du sujet. Nous détectons les points du visage courant  $X_c^i$  avec GenFaceTracker [9]. Ayant les différents coefficients  $\alpha_i, \beta_i$  et  $\gamma_i$  de chacun des 10 points du visage souriant, et connaissant les coordonnées des triangles  $(u_i, v_i, w_i)$  dans lesquels se trouvent ces points, nous déterminons les positions des 10 nouveaux points du visage courant souriant  $X_{cs}^i$  en utilisant l'équation (2). L'utilisation d'une déformation relative par rapport aux points du visage détecté permet de s'affranchir du problème de pose.

$$X_{cs}^i = \alpha_i X_c^{u_i} + \beta_i X_c^{v_i} + \gamma_i X_c^{w_i} \quad (2)$$

Une des originalités de notre méthode est que nous pouvons générer cette expression avec différentes intensités. Une étude

statistique des trajectoires de sourires sur des base de données nous a montré que le sourire est rectiligne. Nous proposons donc modèle linéaire (la Fig. 3) :

$$X_{cs}^i(d) = X_c^i + (X_{cs}^i - X_c^i) * d \quad (3)$$

Où  $d$  est le coefficient de déformation. L'augmentation de ce coefficient augmente l'intensité de sourire et vice versa.

- Si  $d = 0$ , le résultat est un visage  $X_c^i$  inchangé .
- Si  $d = 1$ , le résultat est une expression de joie  $X_{cs}^i$  correspondant à l'intensité de celui qui a été appris.

Pour finir nous appliquons la méthode Moving Least Squares (MLS) [10, 6] pour transformer le visage détecté  $X_c$  à partir des nouveaux points calculés. La MLS rigide est très efficace pour déformer l'image et optimise la distorsion créée sur l'image en temps réel. Compte tenu du temps nécessaire pour effectuer les déformations et les appliquer en temps réel, nous faisons un compromis temps / esthétique comme dans [6] : nous appliquons l'algorithme du MLS sur des grilles autour de chaque œil et de la bouche, et non pas sur chaque pixel de l'image.



FIGURE 3 – Les sourires générés en variant le coefficient  $d$  de 0 à 1 avec un pas de 0.2.

### 3 Expérimentations et résultats

Dans cette section nous comparons notre méthode spécifique à la personne à deux méthodes de l'état de l'art :[6] et [8].

Nous avons utilisé les 3 bases de données CK [12] (88 sujets), Oulu-CASIA [13] (80 sujets) et MMI [14] (32 sujets dont chaque sujet a deux vidéos souriantes) pour nos expérimentations. Pour chaque sujet nous apprenons un modèle spécifique à l'aide de son image neutre qui est la première image de la séquence et son image souriante qui correspond à l'Apex de l'expression. En faisant varier le coefficient  $d$  de l'intensité, nous générons 15 images souriantes pour chaque sujet. Pour la base de données MMI, nous avons eu l'opportunité d'apprendre le modèle à partir d'une vidéo et le tester sur la deuxième vidéo de la même personne (cas applicatif réel).

Pour tester notre méthode, nous avons défini une mesure permettant d'évaluer la distance entre le sourire généré et le sourire réel de la personne. Nous choisissons de calculer les angles entre la trajectoire du sourire réel et une trajectoire générée par une méthode, comme illustré dans la Fig. 4. Ce choix à été réalisé suite à une étude statistique qui montre que le trajectoire sont rectilignes. Ces trajectoires représentent le déplacement des points au cours d'un sourire. Pour déterminer les angles, nous utilisons la régression linéaire  $Y = a_{GT}X + b_{GT}$  de la trajectoire réelle. De même,  $Y = aX + b$  est régressé à partir des points générés avec l'une des trois méthodes. Les angles sont calculés comme suit :

$$\theta = \tan^{-1} \left| \frac{a_{GT} - a}{1 + a_{GT}a} \right| \quad (4)$$

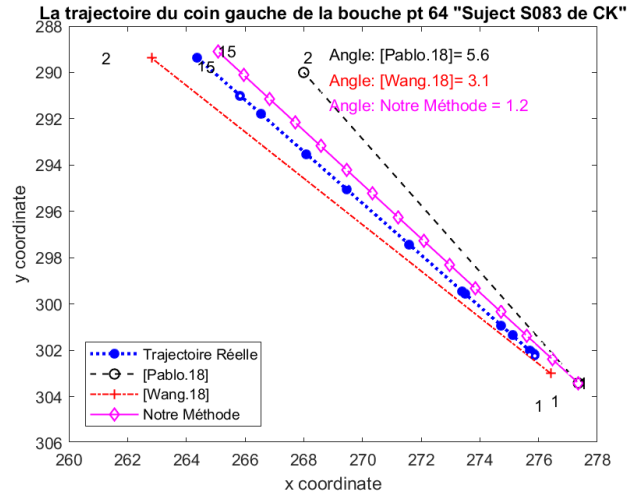


FIGURE 4 – La trajectoire réelle du point de coin gauche de la bouche et les trajectoires générées. Les déplacements de points sont présentés avec une régression linéaire. L'angle entre la trajectoire réelle et la trajectoire de notre méthode est inférieur à l'angle calculé avec [6] et avec [8].

Ayant les angles de chaque trajectoire des points caractéristiques, nous calculons la moyenne  $\bar{\theta}$  et l'écart type  $\sigma_{theta}$  de ces angles sur les 3 bases de données.  $\bar{\theta}$  représente l'erreur entre la trajectoire réelle et une trajectoire générée, elle doit donc être aussi petite que possible.

La table 1 présente les résultats statistiques du point du coin gauche de la bouche. MMI\* représente les résultats trouvés lorsque nous apprenons un modèle sur une vidéo et en le testant sur une deuxième vidéo de la même personne. Les résultats quantitatifs montrent que notre méthode a le plus petit  $\bar{\theta}$  et donc elle génère des trajectoires plus proches de la vérité terrain que les méthodes génériques [6] et [8]. La même étude a été réalisée pour les 10 points impliqués dans l'expression de joie.

TABLE 1 – La moyenne et écart type des angles calculés avec les 3 méthodes sur les 3 bases de données pour le point du coin gauche de la bouche sur tous les sujets.

Méthode	Base de données			
	CK	Oulu-CASIA	MMI	MMI*
[Pablo.18]	$\bar{\theta}= 12.10$ $\sigma_{\theta}= 9.73$	$\bar{\theta}= 17.81$ $\sigma_{\theta}= 15.94$	$\bar{\theta}= 12$ $\sigma_{\theta}= 13.70$	$\bar{\theta}= 12$ $\sigma_{\theta}= 13.70$
[Wang.18]	$\bar{\theta}= 16.16$ $\sigma_{\theta}= 12.24$	$\bar{\theta}= 26$ $\sigma_{\theta}= 18.21$	$\bar{\theta}= 15.26$ $\sigma_{\theta}= 12.66$	$\bar{\theta}= 15.26$ $\sigma_{\theta}= 12.66$
<b>Notre méthode</b>	$\bar{\theta}= 6.65$ $\sigma_{\theta}= 7.56$	$\bar{\theta}= 5.58$ $\sigma_{\theta}= 5.37$	$\bar{\theta}= 4.38$ $\sigma_{\theta}= 5.27$	$\bar{\theta}= 7.50$ $\sigma_{\theta}= 9.68$

La Fig. 5 illustre les résultats obtenus pour une intensité de 1 avec les 3 méthodes. Nous observons qu'avec la méthode géométrique de Pablo et al.[6], le coin des lèvres est systématiquement remonté (pente forte) pour les 3 sujets (a, b, c). Nous remarquons que le sujet (b) a un sourire plat (faible pente) en réalité alors que le sourire généré avec [6] est croissant. En effet, cette méthode applique la même déformation quel que soit le sujet. La fidélité visuelle montre que le sourire généré

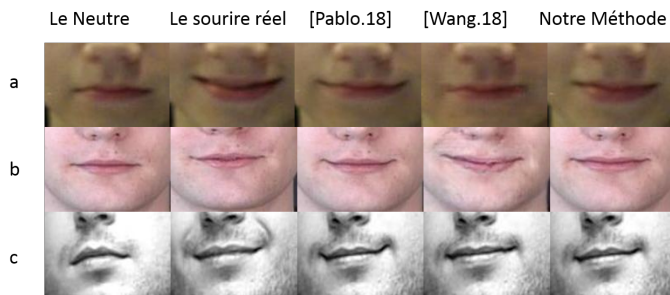


FIGURE 5 – Images souriantes générées avec les différentes méthodes [6], [8] et la notre sur les 3 bases de données CK, MMI et Oulu-CASIA respectivement.

avec notre méthode est plus réel. Les résultats avec le GAN [8] montrent que les sourires générés sont différents pour les 3 sujets. Néanmoins, le GAN génère des sourires réalistes mais pas réels. Pour le sujet (a) nous remarquons que son sourire est croissant alors que le GAN génère un sourire plutôt plat, l'angle entre la trajectoire réelle et la trajectoire générée par le GAN est de 10.41 alors que l'angle est égal à 7.5 avec la trajectoire générée par notre méthode. Pour le sujet (b) le sourire réel est plat mais le GAN génère un sourire un peu croissant. Nous trouvons également que l'angle avec notre méthode est égal à 4.8 alors qu'il est égal à 17.4 avec le GAN. En effet, notre méthode permet de générer l'expression la plus proche de la vérité terrain. Néanmoins, nous ne pouvons pas synthétiser les dents ni les rides (globalement la texture) de la personne puisque le modèle se focalise sur la forme du sourire. Par ailleurs, la déformation est appliquée même si le visage est déjà souriant.

## 4 Conclusion

Dans cet article, nous avons proposé un algorithme capable d'apprendre un modèle spécifique à la personne et de transformer son visage détecté en temps réel afin qu'il apparaisse plus joyeux. Notre méthode permet de synthétiser des expressions propres à la personne avec des intensités réglables. Nos résultats montrent que notre méthode peut générer une expression de joie plus proche de la vérité terrain que deux méthodes récentes de l'état de l'art.

## Références

[1] Tara L Kraft and Sarah D Pressman, “Grin and bear it : The influence of manipulated facial expression on the stress response,” *Psychological science*, vol. 23, no. 11, pp. 1372–1378, 2012.

[2] David T Neal and Tanya L Chartrand, “Embodied emotion perception : amplifying and dampening facial feedback modulates emotion perception accuracy,” *Social Psychological and Personality Science*, vol. 2, no. 6, pp. 673–678, 2011.

[3] Elaine Hatfield, John T Cacioppo, and Richard L Rapson, “Emotional contagion,” *Current directions in psychological science*, vol. 2, no. 3, pp. 96–100, 1993.

[4] Keita Suzuki, Masanori Yokoyama, Shigeo Yoshida, Takayoshi Mochizuki, Tomohiro Yamada, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose, “Faceshare : Mirroring with pseudo-smile enriches video chat communications,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017, pp. 5313–5317.

[5] Shigeo Yoshida, Tomohiro Tanikawa, Sho Sakurai, Michitaka Hirose, and Takuji Narumi, “Manipulation of an emotional experience by real-time deformed facial feedback,” in *Proceedings of the 4th Augmented Human International Conference*. ACM, 2013, pp. 35–42.

[6] Pablo Arias, Catherine Soladie, Oussema Bouafif, Axel Robel, Renaud Segulier, and Jean-Julien Aucouturier, “Realistic transformation of facial and vocal smiles in real-time audiovisual streams,” *IEEE Transactions on Affective Computing*, 2018.

[7] Hui Ding, Kumar Sricharan, and Rama Chellappa, “Exprgan : Facial expression editing with controllable expression intensity,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[8] Xueping Wang, Weixin Li, Guodong Mu, Di Huang, and Yunhong Wang, “Facial expression synthesis by u-net conditional generative adversarial networks,” in *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*. ACM, 2018, pp. 283–290.

[9] Dynamixyz, “Genfacetracker : person-independent real-time face tracker : <http://www.dynamixyz.com>,” 2017.

[10] Scott Schaefer, Travis McPhail, and Joe Warren, “Image deformation using moving least squares,” in *ACM transactions on graphics (TOG)*. ACM, 2006, vol. 25, pp. 533–540.

[11] Paul Ekman and Wallace V Friesen, *Facial Action Coding System : Investigator's Guide*, Consulting Psychologists Press, 1978.

[12] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian, “Comprehensive database for facial expression analysis,” in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 46–53.

[13] Guoying Zhao, Xiaohua Huang, Matti Taini, Stan Z Li, and Matti Pietikäinen, “Facial expression recognition from near-infrared videos,” *Image and Vision Computing*, vol. 29, no. 9, pp. 607–619, 2011.

[14] Michel Valstar and Maja Pantic, “Induced disgust, happiness and surprise : an addition to the mmi facial expression database,” in *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC) : Corpora for Research on Emotion and Affect*, 2010, p. 65.