

Artus : Animation Réaliste par Tatouage audiovisuel à l'Usage des Sourds

P. Bas¹, J. Liénard¹, J.M. Chassery¹, D. Beautemps² et G. Bailly¹

¹ Laboratoire des Images et des Signaux (LIS)
INPG/CNRS
BP 46

F-38402 Saint-Martin-d'Hères CEDEX, France

² Institut de la Communication Parlée (ICP)
46 avenue Felix Viallet
F-38031 Grenoble CEDEX 1, France

Patrick.Bas@lis.inpg.fr, Joel.Lienard@lis.inpg.fr,
Jean-Marc.Chassery@lis.inpg.fr, beautemps@icp.inpg.fr,
bailly@icp.inpg.fr

Artus: Animation Réaliste par Tatouage audiovisuel à l'Usage des Sourds

P. Bas⁽¹⁾, J. Lienard⁽¹⁾, J-M Chassery⁽¹⁾, D. Beautemps⁽²⁾, G. Bailly⁽²⁾

(1): Laboratoire des Images et des Signaux (LIS) INPG/CNRS,

BP. 46 38402 Saint Martin d'Hères cedex

(2): Institut de la Communication Parlée (ICP),

46, avenue Félix Viallet, 38031 GRENOBLE CEDEX 1

Patrick.Bas@lis.inpg.fr,

Joel.Lienard@lis.inpg.fr, Jean-Marc.Chassery@lis.inpg.fr, beautemps@icp.inpg.fr, bailly@icp.inpg.fr

Abstract—Le projet Artus a pour objectif d'insérer dans les émissions télévisées classiques une information imperceptible permettant d'animer à la réception un clone de synthèse. Ce clone 3D communique avec des téléspectateurs sourds grâce à un code manuel appelé Langage Parlé Complété (LPC) venant en appoint de la lecture labiale. Dans le cadre de ce projet, les mouvements du clone sont directement insérés en utilisant des techniques de tatouage. De telles techniques permettent l'insertion d'une information indélébile et imperceptible. Ce projet traite de problèmes scientifiques multiples comme la conversion de texte en LPC, le codage des mouvements du clone, le tatouage haute-capacité ou encore l'animation 3D du clone. Artus est composé de 7 partenaires dont 4 laboratoires universitaires, une chaîne audiovisuelle et 2 partenaires industriels.

I. INTRODUCTION

Actuellement 3 à 5 millions de Français sont sourds ou souffrent d'un déficit auditif ce qui représente 6 à 8% de la population française. La surdité est un handicap auquel beaucoup de personnes sont ou seront confrontées puisqu'il est souvent lié à l'âge : la population des sourds et malentendants est composée de 11% de personnes de moins de 10 ans, de 27% de personnes âgées entre 10 et 60 ans et de 27% de personnes de plus de 60 ans.

Le projet RNRT Artus propose de développer des applications multimédia permettant d'augmenter l'intelligibilité des émissions télévisées vis à vis des personnes sourdes et malentendants. Pour ce faire, un clone, composé d'un visage et d'une main, est animé en superposition de documents télévisés classiques.

La Figure 1 présente un résumé des différentes tâches du projet Artus. Pour effectuer l'animation du clone, les paramètres de mouvements sont synthétisés à partir du télétexte issu du document télévisé. La transmission de ces différents paramètres au sein du flux audio-vidéo s'effectue par techniques de tatouage haute capacité. Une telle approche permet de pouvoir animer le clone quel que soit le format de diffusion (câble, satellite, hertzien). Un terminal situé chez

l'utilisateur est chargé de décoder les informations dissimulées dans l'émission télévisée et d'animer le clone 3D à partir de ces informations.

Ce document a pour objectif de présenter d'une part le contexte scientifique du projet (le langage parlé complété et le tatouage), et d'autre part de souligner les enjeux scientifiques et techniques qui y sont rattachés.

II. PRÉSENTATION DU PROJET

Cette section présente les deux domaines qui sont à l'origine de ce projet, à savoir le Langage Parlé Complété et le tatouage de contenus audiovisuels.

A. Le Langage Parlé Complété

1) *Présentation générale*: Le Langage Parlé Complété est l'adaptation française du Cued Speech[5]. Ce code permet de distinguer les sosies labiaux (groupes de phonèmes qui lors de leur prononciation ont une même forme aux lèvres, également appelés visèmes) et ainsi de lever les ambiguïtés de l'information labiale par la vision. C'est un moyen de "rendre les sons visibles".

Le locuteur exécute, pendant qu'il parle, une succession de codes (clés) à l'aide de la main. Chaque code est composé de deux paramètres : la configuration de la main (position des doigts les uns par rapport aux autres) et la position de la main sur un côté du visage. La forme adoptée par la main lève les ambiguïtés des consonnes tandis que la position où se place la main lève les ambiguïtés des voyelles (cf Fig. 2). Ainsi la main ayant une position donnée pour une voyelle V et présentant une clé digitale pour la consonne C peut coder simultanément la syllabe CV.

Les phonèmes qui peuvent être confondus sur les lèvres (par exemple : /p/, /b/, /m/) sont représentés par des clés différentes afin de bien les différencier, tandis que ceux qui ont des images labiales clairement distinctes (par exemple : /p/, /d/, /j/) sont regroupés sous une même clé. Il en est de même pour les voyelles. Ainsi, on peut transmettre toute l'information phonologique avec un nombre limité de clés.

Cependant, il est important de noter que les configurations et

Projet ARTUS: Animation Réaliste par Tatouage audiovisuel à l'Usage des Sourds



Fig. 1. Présentation générale du projet Artus

les positions de la main ne fournissent pas à elles seules l'information phonologique non ambiguë (elles sont compatibles avec plusieurs phonèmes différents). C'est la combinaison des mouvements des lèvres et de la main qui permet de transmettre visuellement cette information avec la même précision que l'information audio-visuelle pour les entendant [2].

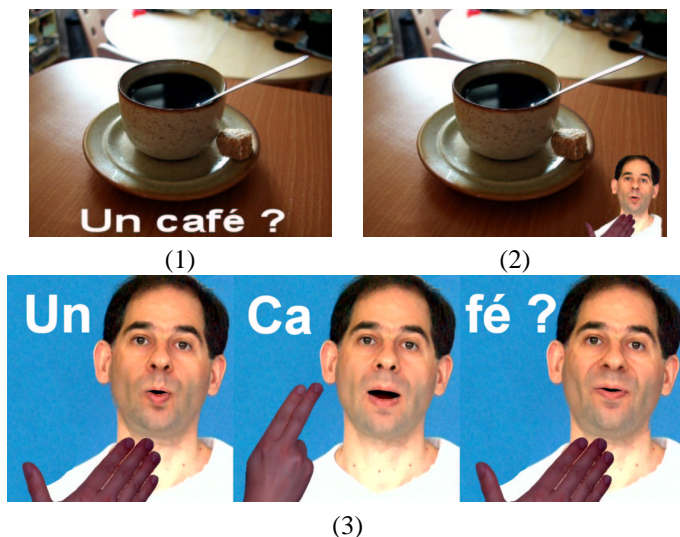


Fig. 3. Emploi du Langage Parlé Complété dans des émissions télévisées : (1) :Séquence avec télétexte, (2) Séquence où le télétexte est remplacé par un clone animé,(3) Séquence de gestes en LPC représentant la formulation "un café"

2) *Avantages du LPC:* L'emploi du Langage Parlé Complété peut être motivé par plusieurs facteurs :

- **l'intégration** : le LPC permet l'intégration des sourds dans la communauté entendant. En effet il permet aux enfants sourds exposés au code dès le plus jeune âge de construire une représentation complète du système phonologique de la langue parlée. Ceci favorise le développement du langage ainsi que l'accès à la lecture et l'écriture de manière comparable à un enfant bien entendant [6].

- **l'apprentissage** : le LPC est un code facile à apprendre car lié aux traits phonétiques utilisés dans la langue, donc limité à une dizaine de contrastes d'articulations non visibles. En quelques mois, les entendants peuvent facilement apprendre à coder, la pratique ne servant ensuite qu'à accroître la vitesse de codage.
- **la gestuelle** : La gestuelle du LPC utilise une seule main dans un espace réduit sur le côté du visage. En termes techniques, l'espace exploité par le LPC est bidimensionnel car il exploite des gestes "plan" d'une main référencée au visage.
- **le cadrage** : D'un point de vue technique, l'étude et la synthèse des gestes du LPC est facile à appréhender. Le LPC permet en effet un cadrage serré sur le visage et donc l'usage d'une petite fenêtre d'insertion sur l'écran.

Ce projet nous permet d'appréhender avec le LPC la transmission et le codage de mouvements liés à la parole. Les solutions développées dans ce cadre permettront d'initier des pistes pour la transmission de gestes plus complexes tels que ceux de la Langue des Signes (LSF) qui font intervenir notamment le buste et les bras dans un espace plus étendu.

B. Le tatouage audio et vidéo

Après l'utilisation du LPC, les techniques de tatouage représentent la deuxième notion clef de ce projet.

Le tatouage peut se définir comme l'insertion d'une information imperceptible et indélébile au sein d'un contenu. La nature de ce contenu peut être multiple, il peut s'agir par exemple d'images, de sons, de vidéos, de textes ou encore de composants électroniques. Dans le cadre du projet Artus, nous nous attacherons à tatouer des contenus audio-vidéo représentés par les émissions télévisées. L'information insérée sera ici les mouvements d'un clone communicant en Langage Parlé Complété.

Initialement, les techniques de tatouage ont été conçues afin de protéger des oeuvres numériques grâce à l'insertion d'une information renseignant sur l'identité du propriétaire ou de

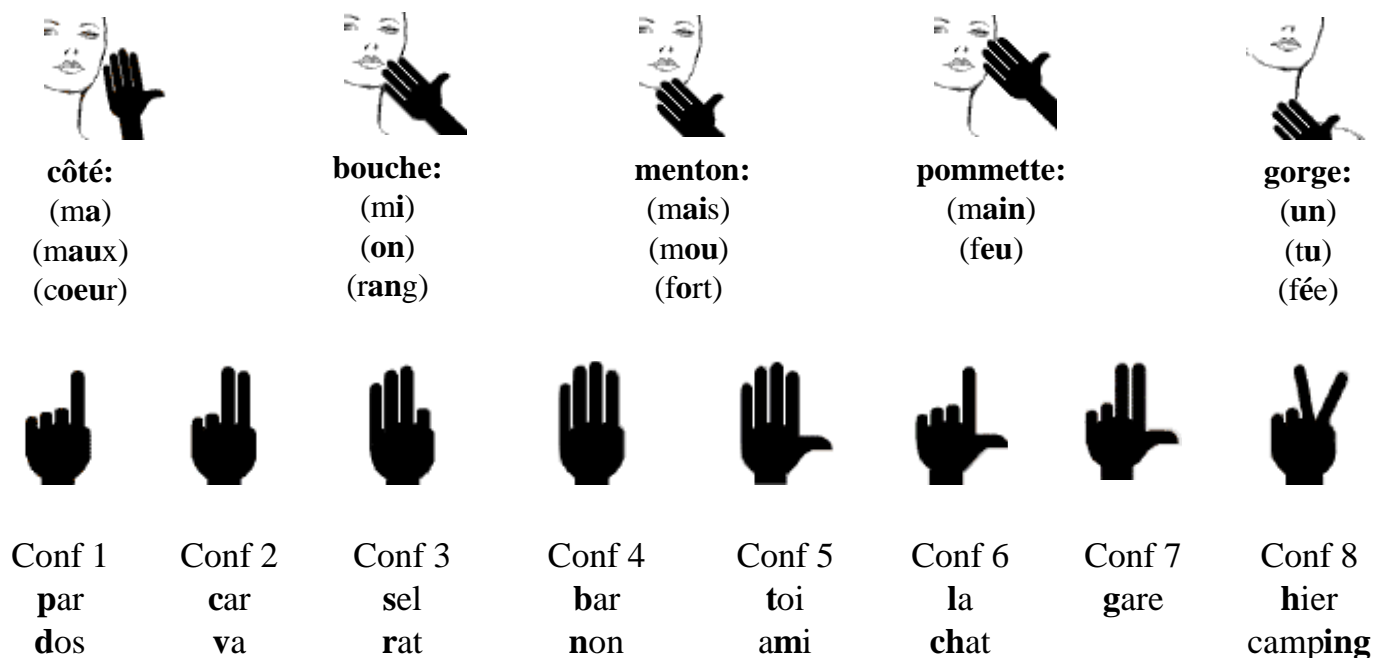


Fig. 2. Les différentes positions de la main pour coder les voyelles (en haut) et les consonnes (en bas).

l'acheteur du contenu. Ces techniques cherchant à protéger les droits des auteurs s'avèrent cependant extrêmement complexes à mettre en oeuvre car elles doivent faire face à des attaques directement destinées à effacer le tatouage.

Dans le cadre du projet Artus, l'utilisation du tatouage a été envisagée pour permettre d'étendre un contenu multimédia en ajoutant d'une fonctionnalité initialement non prévue lors de la création du format initial. Le tatouage permet dans ce projet le transport des mouvements du clone dans des émissions télévisées, quelque soit le mode de diffusion (hertzien, numérique, ou par câble) ou l'opérateur de diffusion (CanalSat, TPS, ...). Dans la littérature, plusieurs autres exemples utilisent le tatouage pour étendre des contenus, on peut citer l'utilisation du tatouage pour le codage d'images couleurs [4], la création d'images stéréoscopiques étendues [3], la correction d'erreurs pour la transmission de séquences vidéo [7], l'utilisation du tatouage pour faciliter la fouille de données, ou encore l'insertion de code barres invisibles pour servir à des fins publicitaires [1].

III. PLUSIEURS VERROUS TECHNOLOGIQUES

Ce projet offre un enjeu applicatif de grande envergure tout en soulevant des problèmes scientifiques extrêmement motivant pour les différents partenaires de ce consortium. L'objectif de cette section est de présenter les tâches et les enjeux scientifiques clés de ce projet.

A. Transduction du télétexte en Langage Parlé Complété

Dans le cadre d'Artus, nous proposons de générer le LPC directement à partir du télétexte contenu dans la plupart des émissions diffusées. Cette tâche peut être décomposée en deux

niveaux. Dans un premier temps, un signal de parole doit être généré à partir du télétexte. Ce signal doit ensuite être analysé pour générer les différents paramètres nécessaires à la compréhension du LPC : la forme des lèvres, la forme et la position de la main et la manière dont s'effectuent les transitions.

B. Modélisation et codage des mouvements du Langage Parlé Complété

Pour que les paramètres nécessaires à l'animation du clone puissent être transmis par tatouage, il faut que leur représentation soit la plus condensée possible. Il est donc nécessaire de procéder à un codage source de ces paramètres. De par le caractère multimodal de cette information (positions des lèvres, de la main et transitions entre les mouvements), la dépendance entre ces paramètres et l'évolution des paramètres dans le temps doit être analysée puis ensuite modélisée. Ces modèles, linéaires ou non, permettront de réduire le nombre de paramètres à transmettre tout en ne dégradant pas au final l'intelligibilité du clone communicant (cf. Fig. 4). Cette tâche a débuté par l'acquisition de données réelles des mouvements d'une codeuse LPC qui ont été ensuite utilisées afin de concevoir des modèles de mouvements.

C. Développement d'algorithmes de tatouage audio-vidéo

Concernant le tatouage des contenus audiovisuels, les contraintes spécifiques à cette application sont multiples :

- L'information ajoutée doit être aussi importante que possible de manière à ce que l'animation du clone soit aussi réaliste que possible. Il faut donc concevoir des algorithmes de tatouage haute capacité (cf Fig. 5).
- Le décodage de l'information dissimulée doit être possible quelque soit la chaîne de diffusion. Ainsi le message doit être lisible après une compression du signal, une distorsion

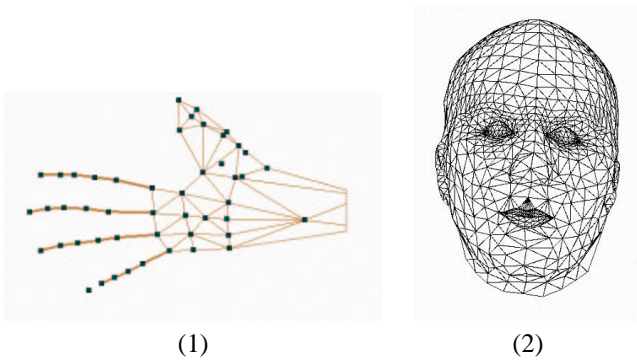


Fig. 4. Maillages utilisés pour analyser et modéliser les mouvements du codeur LCP. (1) :Maillage de la main, (2) : Maillage du clone.

due à une réception de faible qualité ou encore transcodage du flux audio-vidéo.

- Les techniques de tatouage mises en oeuvre doivent également être de complexité très faible afin de pouvoir être utilisées pour un décodage en temps-réel. La complexité doit être inférieure à la complexité nécessaire au décodage de la source vidéo.
- La nature des contenus doit également être prise en compte, par exemple le mouvement d'objets dans une séquence augmente la visibilité du tatouage : une marque "fixe" ajoutée sur un objet qui se déplace est d'avantage perceptible que si l'objet reste fixe.
- L'information à dissimuler doit être répartie entre deux contenus de natures totalement différentes : le son et la séquence vidéo. Afin d'offrir plus de souplesse dans le fonctionnement du codeur et du décodeur, le débit d'insertion doit également être variable pour chacune des sources.

D. Animation réaliste du clone

Le terminal client doit être capable de décoder le tatouage en temps réel mais aussi d'effectuer le rendu d'un clone aussi réaliste que possible. Cette dernière condition assure une intelligibilité de bonne qualité. La société Attitude-Studio, partenaire du projet Artus, est spécialisée dans la génération de clones extrêmement réalistes (cf Fig 6). Elle a en charge de créer un moteur d'animation temps-réel spécifique au décodage du clone d'Artus.

IV. PRÉSENTATION DES PARTENAIRES

Le consortium d'Artus est composé de deux partenaires industriels (Thalès et Attitude-Studio), un partenaire audiovisuel public (Arte-France) et 4 partenaires universitaires (LIS/INPG, ICP/INPG, TSI/ENST et HEUDIASYC/UTC). Cette section présente le rôle de chaque partenaire dans ce projet.

A. Thalès

Thalès Communications -THC- (Gennevilliers) a un très large spectre d'activités dans le domaine des communications couvrant les équipements de radio mobiles, les grands réseaux terrestres, les communications aéronautiques, spatiales et navales.

Dans le cadre d'Artus, Thalès, associée à sa spin-off Nextamp,



Fig. 6. Exemple d'un clone réaliste conçu par Attitude-Studio.

pour objectif de développer un tatoueur vidéo temps réel et de concevoir le terminal cible permettant le décodage du tatouage.

B. Attitude Studio

ATTITUDE STUDIO est le premier prestataire spécialisé dans la conception, la fabrication et l'exploitation de personnages virtuels haut de gamme.

Ses missions au sein du projet sont d'aider à la modélisation d'un clone communiquant en LPC ainsi qu'à la conception du moteur 3D permettant l'animation de ce clone.

C. Arte-France

ARTE France - issue de La SEPT, préfiguration d'une chaîne culturelle à vocation européenne - est chargée de produire, de coproduire et d'acheter des programmes pour ARTE.

Son rôle au sein du projet Artus est d'une part de réaliser le lien entre le monde de l'audiovisuel et celui des personnes malentendantes et, d'autre part de mettre à la disposition du consortium des moyens permettant de tester les techniques développées.

D. LIS/INPG

Le LIS (Laboratoire des Images et des Signaux) est un laboratoire de l'Institut National Polytechnique de Grenoble (INPG), Unité Mixte de Recherche du CNRS (UMR 5083). Il est spécialisé en traitement du signal et en traitement de l'image.

Le LIS est également chef de file du projet Artus. Outre la coordination du projet, ce laboratoire est impliqué dans le développement de tatoueurs audio et vidéo haute capacité.

E. ICP/INPG

L'ICP (Institut de la Communication Parlée) est un laboratoire associé au CNRS (UMR n°5009) et à deux universités de Grenoble : l'INPG et l'Université Stendhal. L'ICP développe ses recherches à la fois en Sciences de l'Ingénieur et en Sciences Humaines et étudie la parole dans tous ses états : signaux, langage et cognition.

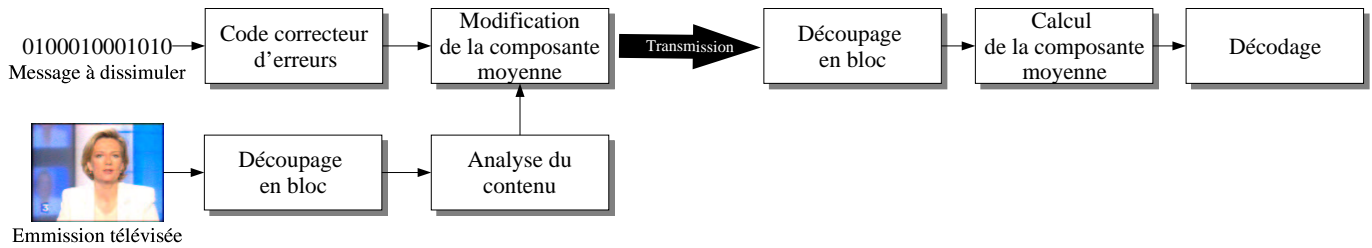


Fig. 5. Principe d'un schéma de tatouage haut-débit vidéo basé sur un codage par quantification indexée.

Le rôle de l'ICP dans ce projet est de fournir des techniques permettant d'acquérir, de synthétiser et modéliser le Langage Parlé Complété.

F. TSI/ENST

Le laboratoire TSI de l'Ecole Nationale Supérieure des Télécommunications à Paris possède une compétence dans les aspects théoriques des communications numériques, du traitement du signal, de la compression de la parole et des images, ainsi que dans les aspects théoriques du traitement des images. Dans ce projet, le laboratoire TSI a pour objectif de concevoir et modéliser des algorithmes de tatouage audio et vidéo.

G. HEUDIASYC/UTC

Le laboratoire Heudiasyc/UTC (UMR CNRS 6599) de l'Université de Technologie de Compiègne est spécialisé dans les domaines de l'analyse de données, de la reconnaissance des formes pour le diagnostic de systèmes et, du traitement du signal et des images.

Ce laboratoire participe à l'étude de la représentation des expressions faciales du codeur LPC ainsi qu'au développement de tatoueurs vidéo.

V. CONCLUSION

Le projet Artus, dont la durée totale est de trois ans, est maintenant à mi-parcours. Les différentes tâches sont en cours de réalisation aussi bien sur la partie concernant le Langage Parlé Complété, que sur la partie tatouage. Ce projet a également permis d'impliquer trois doctorants ainsi que de nombreux étudiants. La deuxième phase de ce projet se focalise actuellement sur l'implémentation et les tests de la chaîne de traitements proposés dans Artus.

REFERENCES

- [1] A. M. Alattar. Smart images using digimarc corporation's watermarking technologie. In *Proc. SPIE*, pages 264–273, January 2000.
- [2] V. Attina, D. Beautemps, M.-A. Cathiard, and M. Odisio. Toward an audiovisual synthesize for cued speech : Rules for cv french syllables. In *proceedings of AVSP 2003*, to appear.
- [3] P. Barralon, P. Bas, and J.-M. Chassery. Projet de fin d'étude : Création d'images stéréoscopiques étendues par techniques de tatouage. Master's thesis, INPG, 2002.
- [4] P. Campisi, D. Kundur, D. Hatzinakos, and A. Neri. Compressive data-hiding : An unconventional approach for improved color image coding. *Applied Signal Processing*, 2002(3), Feb 2002.
- [5] O. Cornett. Cued speech. *American Annals of the Deaf*, 112 :3–13, 1967.
- [6] J. Leybaert. Phonology acquired through the eyes and spelling in deaf children. *Journal of Experimental Child Psychology*, 75 :291–318, 2000.
- [7] Robie and Mersereau. Video error correction using steganography. *Applied Signal Processing*, 2002(3), Feb 2002.