CNRS INFO • N° 397
OCTOBRE-NOVEMBRE 2001

LES CLONES ONT LA PAROLE !

Lecture labiale et clones parlants

La parole est un ensemble de gestes rendus visibles et audibles. La construction d'un clone virtuel capable de produire du son à partir de gestes permet d'étudier les relations entre programmation motrice et perception visuelle et acoustique. Les chercheurs de l'Institut de communication parlée[1] (ICP) ont choisi d'étudier un locuteur réel dont les gestes de parole sont enregistrés par divers capteurs (mouvements visibles par vidéo, invisibles par cinéradiographie, articulographie, imagerie par résonance magnétique...). Ces études du mouvement *in vivo* ont fait apparaître les « degrés de liberté » des articulateurs de la parole, ces mouvements élémentaires que les organes peuvent exécuter indépendamment les uns des autres pour mettre en forme la géométrie du conduit vocal.

[1] *CNRS-INPG Grenoble-Université Stendhal.*

Malgré la complexité de la musculature, les chercheurs de l'ICP ont montré qu'une dizaine de degrés de liberté étaient suffisants pour décrire l'ensemble des mouvements de parole des locuteurs étudiés : deux pour la mâchoire (ouverture/fermeture et avancée/rétraction), trois pour les lèvres (étirement/protrusion, ouverture/fermeture, élévation/abaissement), un pour le vélum (ouverture/fermeture), un pour le larynx (élévation/abaissement) et quatre pour la langue (avant/arrière, tendue/comprimée, élévation/abaissement et avancée/rétraction de la pointe de la langue).

Quatre des six gestes faciaux élémentaires impliqués par les mouvements de parole. (a) fermeture/ouverture de la mâchoire ; (b) étirement/arrondissement des lèvres ; (c) ouverture/fermeture des lèvres ; (d) abaissement/élévation des lèvres. © CNRS ICP.

Les chercheurs ont utilisé des clones articulés qui ont l'apparence du locuteur original. Ils constituent de véritables outils de recherche sur le développement du langage chez l'enfant, l'émergence du langage chez l'homme ou la structure des langues du monde. Ils trouvent leurs applications dans différents domaines :

• *Les télécommunications virtuelles*

La visiophonie (transmission simultanée du son et de l'image) y est remplacée par un système de labiophonie où les interlocuteurs sont représentés par leurs clones respectifs, animés par les mouvements articulatoires estimés. Le clone participe à des téléconférences virtuelles.

© CNRS ICP

On suit les mouvements faciaux en estimant une projection optimale du modèle articulatoire sur l'image (issue ici d'une microcaméra placée à T2 cm du visage). Les paramètres estimés peuvent alors animer à distance le clone du locuteur (ou celui d'un autre par une procédure d'encodage/recodage des paramètres articulatoires propre à chaque clone dans le format MPEG4/SNHC). Notez ici l'inférence des mouvements de profil à partir d'une vue de face.

CNRS INFO • N° 397
OCTOBRE-NOVEMBRE 2001

• *La réalité augmentée*

Les mouvements visibles et invisibles des clones estimés pour un apprenant ou un patient peuvent les guider dans des tâches d'apprentissage, de rééducation et aider l'orthophoniste ou le chirurgien à identifier des déficiences ou des pathologies. La peau du clone virtuel est rendue partiellement transparente et l'on peut voir ainsi les mouvements des articulateurs invisibles de la parole.

• *Les technologies vocales (synthèse et reconnaissance de parole)*

Ces technologies bénéficient d'une intégration forte des composantes acoustiques et visuelles : le message synthétique gagne en intelligibilité, en réalisme et en crédibilité alors que la compréhension de la parole naturelle gagne en robustesse.

La parole est donc un moyen de communication multimodal : audible, visible, voire « touchable »[2]. Les clones parlants de l'ICP laissent entrevoir des applications de télétransportation du corps, permettant de communiquer plus efficacement et à moindre coût et de simuler des interventions sur un organe pas assez agile, empêché ou malade. Les têtes parlantes sont aussi un outil de recherche permettant de s'interroger sur les liens existant entre les systèmes de production et de perception de parole, le langage et la cognition, thème central des recherches fondamentales menées à l'ICP.

Références :

• Elisei, F., M. Odisio, G. Bailly, and P. Badin. (2001) Creating and controlling video-realistic talking heads. In *Auditory-Visual Speech Processing Workshop*. Scheelsminde, Denmark.

• Revèret, L., G. Bailly, and P. Badin. (2000) MOTHER: a new generation of talking heads providing a flexible articulatory control for video-realistic speech animation. In *Proceedings of the International Conference on Speech and Language Processing*. Beijing, China. pp. 755-758.

La lecture labiale

La lecture labiale (le fait de lire sur les lèvres) est très pratiquée par les sourds qui peuvent ainsi reconnaître 40 à 60 % des phonèmes et 10 à 20 % des mots. La compréhension par lecture labiale est imparfaite car tous les gestes articulatoires qui permettent de produire des sons ne sont pas visibles de « l'extérieur » du conduit vocal.

La lecture labiale ne sert pas qu'aux malentendants : chacun a la capacité de lire sur les lèvres et peut s'en servir notamment en situation de communication bruitée. L'absence d'image lors d'une conversation téléphonique complique la compréhension d'une langue étrangère : la vision joue donc un rôle essentiel dans le langage.

L'étude des relations sensori-motrices en parole est complexe à double titre. Le conduit vocal qui sert à produire les sons est mis en forme par de nombreux organes visibles (lèvres, joues…), partiellement visibles (langue, mâchoire,…), voire invisibles bien que parfaitement audibles (vélum, larynx…). La position de ces organes est contrôlée par une centaine de muscles qui doivent être coordonnés de manière très précise. Les relations entre la géométrie et le son produit sont régies par les lois de l'acoustique et de très faibles changements de forme peuvent induire de grands changements acoustiques notamment de mode d'écoulement : ainsi, par une variation de quelques mm² de l'aire entre les lèvres, on peut passer de la voyelle « ou » à la fricative « f » ou à l'occlusive « p ».

Une illustration spectaculaire de l'importance de la perception visuelle de la parole est illustrée par une illusion célèbre, « l'effet McGurk ». Dans une première expérience, on projette sur un écran le visage en gros plan d'un locuteur ou d'une locutrice qui prononce à plusieurs reprises la séquence : « A-di ». C'est du moins ce que nous croyons entendre. Mais lorsque l'on écoute la même séquence sans image, nous n'entendons plus « A-da », mais « A-ba ». En réalité, l'illusion consiste à faire le montage d'un son de « A-ba » sur une image de… « A-ga » ! L'on ne perçoit ni ce que « dit » l'oreille (« A-ba »), ni ce que « dit » l'œil (« A-ga »), mais un produit de « fusion », « A-da » !

La peau transparente du clone rend visibles les mouvements des organes normalement invisibles de la parole (langue, mâchoire). On se situe ici au centre de la réalisation du /t/ dans la syllabe /ta/. © CNRS ICP

[2] *La méthode TADOMA, mise au point à la fin du XIXe siècle pour enseigner la compréhension et la production de la parole à des enfants sourds-aveugles, consiste à placer la main en contact avec le visage de l'interlocuteur (lèvres, joue, larynx, mandibule), de façon à « sentir » ses gestes vocaux. Les utilisateurs de cette méthode réussissent ainsi à avoir des performances d'intelligibilité remarquables leur permettant de suivre une conversation pratiquement normale.*

Contacts chercheurs :
Institut de la communication parlée (ICP), CNRS-INP Grenoble-Université Grenoble 3,
• Gérard BAILLY,
tél. : 04 76 57 47 11
mél : bailly@icp.inpg.fr

• Jean-Luc SCHWARTZ,
tél. : 04 76 57 47 12
mél : schwartz@icp.inpg.fr

• Pierre BADIN,
tél. : 04 76 57 48 26
mél : badin@icp.inpg.fr

Contact département des Sciences et technologies de l'information et de la communication du CNRS :
Armelle TOULEMONDE,
tél. : 01 44 96 53 88
mél : Armelle.Toulemonde@cnrs-dir.fr