Le vrai visage de la parole

La parole anime le visage selon six mouvements élémentaires.

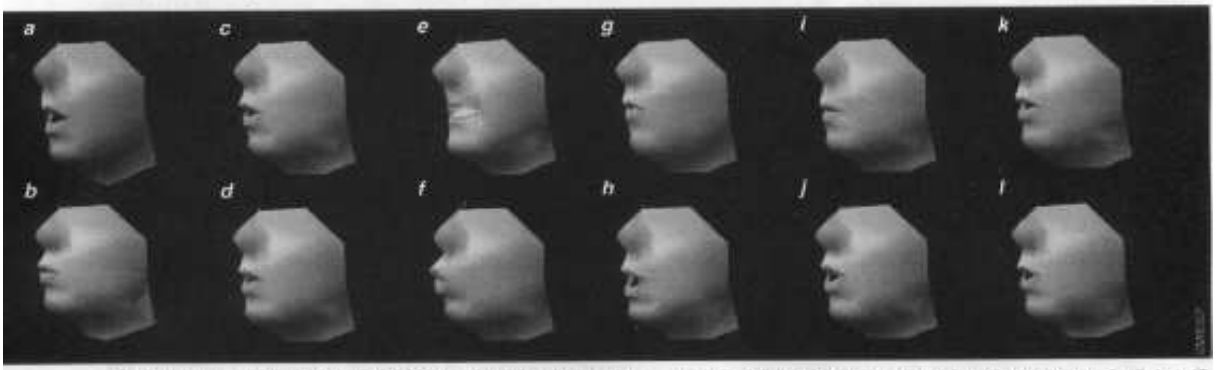
De combien de ficelles usait Geppeto pour donner la parole à Pinocchio? Parmi l'ensemble des muscles du corps humain, un tiers participent à l'animation du visage durant la parole, mais s'imagine-t-on Geppeto manipulant plusieurs centaines de fils simultanément? Non, car d'après une étude de Gérard Bailly et ses collègues de l'Institut de communication parlée de l'Université de Grenoble, le visage est une marionnette nettement moins complexe qu'on ne l'imagine. Cette équipe cherche à comprendre les mécanismes physiologiques de la parole et à les reproduire sous la forme d'un «clone virtuel», une image tridimensionnelle du visage du locuteur.

Le visage est un acteur important de la parole : les lèvres s'ouvrent et se ferment, la mâchoire bascule, le larynx s'abaisse... Ces mouvements sont autant de signes extérieurs de la parole que les sourds apprennent à reconnaître pour pallier leur handicap. En fait, chacun exploite inconsciemment ces indices visuels. Ainsi, dans un environnement bruyant, on utilise des indices visuels pour mieux saisir les conversations : c'est l'effet «cocktail party». En fait, certaines informations apportées par la voix et par les mouvements du visage sont redondantes. De même, on a constaté que l'enseignement des langues étrangères est plus efficace quand il est prodigué par un professeur que lorsque l'élève n'a accès qu'à des enregistrements sonores. Grâce à un clone virtuel qui reproduirait de façon réaliste les mouvements du visage durant la parole, on apprendrait plus aisément les langues, maternelle ou étrangères.

Lorsque nous parlons, plus de 50 muscles interviennent. Pour réaliser un clone virtuel, certaines équipes ont tenté de modéliser chacun de ces muscles par un ressort, mais le résultat n'était pas naturel. Les films d'animation en images de synthèse, qui sont aujourd'hui assez réalistes quant aux visages, adoptent une technique différente : les voix sont décomposées en un certain nombre de phonèmes élémentaires auxquels sont associées des configurations stéréotypées de la bouche ; un logiciel calcule les images intermédiaires entre deux phonèmes, et un animateur finalise le résultat en retouchant les expressions. Une autre méthode consiste à réaliser une «capture de mouvements» : on place des billes sur une personne, on enregistre les coordonnées de ces marqueurs et, par analyse d'images, on reconstruit les mouvements de la personne. C'est une méthode classique dans l'animation numérique qui permet d'obtenir des mouvements naturels. Coûteuse, elle est principalement utilisée pour l'animation des corps.

G. Bailly et ses collègues l'ont appliquée aux mouvements du visage. Ils ont placé 300 billes sur les visages de quatre personnes parlant différentes langues (français, allemand, anglais et arabe), aux endroits des principaux muscles ; chaque bille était repérée par ses trois coordonnées d'espace : 900 coordonnées sont nécessaires pour définir un état du visage. Ils ont ensuite demandé aux personnes de prononcer une centaine de phonèmes différents, allant de voyelles simples à des syllabes plus complexes. L'analyse statistique de l'évolution des 900 coordonnées montre que ces mouvements sont loin d'être indépendants les uns des autres. Seuls six paramètres suffisent à décrire 96 pour cent de la palette des mouvements : deux pour les mouvements de la mâchoire (ouverture/fermeture, avancée/rétraction), trois pour les lèvres (étirement/rétraction, élévation/abaissement pour chacune des deux lèvres) et un pour la gorge (montée/descente du larynx). Les quelque quatre pour cent restant concernent des détails dans les expressions. D'autres expériences ont révélé que l'on peut aussi décrire simplement les mouvements de la langue (commandée par 21 muscles) par quatre paramètres : avancée et rétraction de la pointe ; avancée et rétraction de la langue ; tension et compression ; élévation et abaissement.

L'une des conclusions surprenantes de l'étude est l'absence de différenciation entre les quatre locuteurs étudiés et, par conséquence, entre les quatre langues. D'un sujet à un autre, les mouvements de la bouche et du reste du visage diffèrent, mais les dix mouvements indépendants (les six du visage et les quatre de la langue) sont les mêmes. Ainsi, on peut avancer que les hommes exploitent les mêmes mouvements élémentaires de leur organe vocal pour parler et la spécificité d'un son ne repose que sur une modulation de l'amplitude de ces mouvements. Outre l'italien, Pinocchio aurait pu parler le russe, le finnois ou encore le sanscrit.



Parmi toutes les combinaisons possibles des 50 muscles du visage, seuls six interviennent dans les mouvements utilisés pour parler : l'ouverture *(a)* et la fermeture de la mâchoire *(b)*, la rétraction *(c)* et l'avancée de la mâchoire *(d)*, la rétraction *(e)* et l'étirement des lèvres *(f)*, le relèvement et l'abaissement des lèvres inférieure *(g et h)*, et supérieure *(i et j)*, et la montée *(k)* et l'abaissement du larynx *(l)*.