Département Parole & Cognition
Equipe Machines Parlantes, Agents communicants & Interaction Face-à-face

Sujet MASTER

Asynchronie audiovisuelle. Données expérimentales et modélisation

Contexte

Ce travail s'inscrit dans le cadre des travaux de l'équipe MPACIF sur la synthèse de parole audiovisuelle. L'enjeu de ce travail est de déterminer la tolérance perceptive en cas d'asynchronie entre évènements acoustiques et visuels dans le cas où cette asynchronie affecte un et un seul événement dans la séquence audiovisuelle, les autres étant parfaitement synchrones.
En effet, la plupart des données disponibles dans la littérature concernent l'asynchronie des signaux dans leur totalité (sensibilité au décalage du par exemple au retard de décodage numérique entre les deux canaux) : ces travaux montrent que les téléspectateurs sont plus sensibles à un retard de la vidéo sur l'audio que l'inverse et que cette tolérance est assez importante (entre -40ms et 250ms, dixon et al, 1980). En complément, il s'agit ici d'étudier la tolérance au décalage audiovisuel d'un seul son dans une séquence audiovisuelle par ailleurs parfaitement synchronisée (cf. figure ci-dessous).

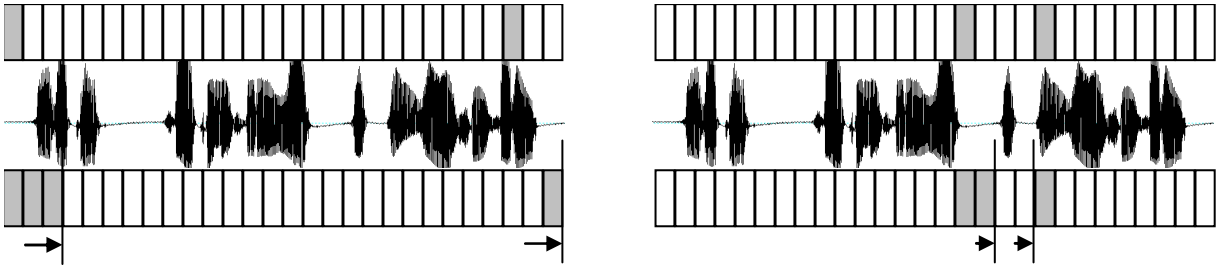

Figure 1. Asynchronie audiovisuelle. A gauche, la bande vidéo est classiquement retardée ou avancée globalement d'une certaine durée par rapport à la bande son. A droite, nous étudions la sensibilité au décalage d'un son élémentaire dans le message. Ici on manipule la bande vidéo afin d'éviter une manipulation plus complexe du signal acoustique.

Sujet

L'idée générale de cette recherche est de montrer expérimentalement que les sujets sont plus sensibles à l'asynchronie entre segments audiovisuels saillants (plosive bilabiale) qu'entre segments moins marqués visuellement ou acoustiquement.
Le travail projeté comporte deux phases :

1. Une phase d'étude des seuils de détection d'asynchronie de segments produits dans des phrases porteuses de type « c'est V1CV2 ça ? » où l'asynchronie audiovisuelle du segment C est manipulée en écourtant/allongeant les durées de V1 et V2. Une étude préliminaire déterminera dans quel espace (son ou séquence d'images) la manipulation unimodale est la moins perceptible
2. Une phase de modélisation de la détection d'asynchronie utilisant un calcul d'intercorrélation entre évènements audiovisuels saillants avec une application à la détection de doublage (qualité du doublage dans les films post-synchronisés, détection d'imposture dans les systèmes de vérification audiovisuelle du locuteur). On utilisera pour ceci des vidéos originales et postsynchronisées par d'autres locuteurs en utilisant la technique de « close-shadowing ».

Environnement de travail

Matlab, Praat

Collaborations

Ce travail s'effectue dans le cadre du PPF « Interactions Multimodales ». Le candidat s'insèrera dans une équipe de six personnes.

Responsables

Gérard BAILLY GIPSA-lab 04 76 57 47 11 Gerard.Bailly@gipsa-lab.grenoble-inp.fr

Références

- Bailly, G., O. Govokhina, F. Elisei and G. Breton (in print). "Lip-synching using speaker-specific articulation, shape and appearance models." Journal of Acoustics, Speech and Music Processing.
- Govokhina, O., G. Breton and G. Bailly (2008). Procédé de repositionnement des frontières de phonèmes pour la synthèse visuelle des mouvements faciaux liés à la parole. INPI - Patent n°FR0757063. France.
- Bailly, G., O. Govokhina, G. Breton, F. Elisei and C. Savariaux (2008). The trainable trajectory formation model TD-HMM parameterized for the LIPS 2008 challenge. Interspeech, Brisbane, Australia, 2318-2321.
- Dixon, N. F. and L. Spitz (1980). "The detection of audiovisual desynchrony." Perception **9**: 719-721.

Gratification

- 398,13 € par mois