

Fiche de projet tutoré / Project form

Evaluation of a multilingual animated speech system

Encadrement / Supervisors

Encadrant: Slim Ouni

Equipe et laboratoire: Multispeech, LORIA

Contact: slim.ouni@loria.fr

Description / Description

1. projet global/global project

In naturally produced speech, the acoustic signal is the result of the deformation of the vocal tract (the jaw, lips and tongue). Indeed, speech communication is essentially multimodal, consisting of the acoustic signal that carries the auditory modality, the movements of the tongue, jaw and lips are the articulatory modality, and the facial deformation represents the visual modality.

In the field of audiovisual speech, we are developing talking heads to study the mechanism of audiovisual speech communication. We recall that a talking head is an animated virtual 3D face of a human speaker synchronously with audio. We have developed an original method to synchronize the movement of lips of 3D face with a given speech audio [BFO19]. Our system can perform lip synchronization in 3 different languages: French, English and German.

Some researchers argue that it is possible to speech animate a 3D face using one unique system independently of the target language [TKY17]. Thus, if such a system has been developed for French, they consider that it is not necessary to build a system for English or any other language, as it is possible to use the French system to animate the 3D face for English, for instance. This speculation means that coarticulation is independent of the language.

2. biblio. UE 705 (semestre 7)

A selection of research papers will be analyzed mainly to study coarticulation, its effects, and if we can consider coarticulation universal or different from one language to another.

3. réalisation. UE 805 (semestre 8)

The purpose of this project is to verify, validate or invalidate the above assumption [TKY17]. To do this, we propose to adapt our multilingual lip synchronization system, to use any of the three built-in languages to speech animate a 3D face in a given target language. The set of phonemes of a given language will be mapped into the target language phonemes. The

mapping will be based on the analysis of the articulatory similarities of the phonemes of two different languages. The second task is to evaluate perceptually the quality of speech animation. A series of perceptual experiments will be set up and the results will be analyzed.

Informations diverses : matériel nécessaire, contexte de réalisation /

Various information: material, context of realization

(see main description)

Livrables et échéancier / Deliverable and schedule

(see main description)

Bibliographie /References (max. 4-5)

[BDO19] Biasutto, T., Dahmani, S., & Ouni, S. (2019, September). Modeling Labial Coarticulation with Bidirectional Gated Recurrent Networks and Transfer Learning. *Processing*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

[TKY17] S.Taylor, T.Kim, Y.Yue, M.Mahler, J.Krahe, A.G.Rodriguez, J. Hodgins, and I. Matthews, "A deep learning approach for generalized speech animation," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 93:1–93:11, Jul. 2017.