The goal of this thesis is to develop a gesture model for a credible communicative robot behavior during speech. The generation of gestures will be studied when the robot is a speaker and when it is a listener. In the context of this thesis, the robot will be replaced by an embodied virtual agent. This allows applying of the outcome of this work in both virtual and real world. It is possible to test the results of this work on a real robot by transferring the virtual agent behavior to the robot, when possible, but it is not an end in itself.

In this thesis, two main topics will be addressed: (1) the prediction of communication-related gesture realization and timing from speech, and (2) the generation of the appropriate gestures during speech synthesis. When the virtual agent is listening to a human interlocutor, the head movement is an important communicative gesture that may give the impression that the virtual agent understands what is said to it and that may make the interaction with the agent more effective. One challenge is to extract from speech, both acoustic and linguistic cues (RADA), to characterize the pronounced utterance and to predict the right gesture to generate (head posture, facial expressions and eye gaze [KCD04]). Synchronizing the gestures with the interlocutor speech is critical. In fact, any desynchronization may induce an ambiguity in the understanding of the reaction of the virtual agent. The gesture timing correlated with speech will be studied. In this work, generating the appropriate gesture during speech synthesis, mainly head posture, facial expressions and eye gaze, will be addressed.

To achieve these goals, motion capture data during uttered speech will be acquired synchronously with the acoustic signal. Different contexts will be considered to achieve the collection of sufficiently rich data. This data will be used to identify suitable features to be integrated within the framework of machine learning techniques. As the data is multimodal (acoustic, visual, gestures), each component will be used efficiently in collecting complementary data. The speech signal will be used in the context of a speech-recognition system to extract the linguistic information, and acoustic features helps to extract non-linguistic information, as F0 for instance. The correlation between gestures and speech signal will also be studied. The aim of the different analyses is to contribute to the understanding of the mechanism of oral communication combined with gestures and to develop a...
model that can predict the generation of gestures in the contexts of speaking and listening.

References


Compétences

Required qualifications

Master of computer science. Good background in modeling, data analysis and machine learning. First experience in speech recognition or in using a deep learning technique will be appreciated.

Language

French or English.

Avantages

Subsidized meals
Partial reimbursement of public transport costs
Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
Professional equipment available (videoconferencing, loan of computer equipment, etc.)
Social, cultural and sports events and activities
Access to vocational training
Social security coverage

Rémunération

Monthly salary after taxes: around 1596,05€ for 1st and 2nd year. 1678,99€ for 3rd year. (medical insurance included).