Introduction

Human speakers make constant adjustments to their vocal output in response to a range of different factors. Local and global changes are made as a function of interferences and noise in the acoustic environment, their familiarity with the listener, the formality of the situation, and the content of the material that is being spoken. Such behavior is not exhibited by state-of-the-art speech synthesizers, and hence, their expressive and communicative abilities are limited by the static nature of their generation process. This project aims to research the possibilities for introducing a feedback path into automatic speech generation such that adjustments can be made as a function of its perceived effectiveness. In order to identify such feedback, this study requires a deep analysis of physiologic and psychological mechanisms involved in human speech production. Several control theories are considered and some initial experiments based on an auditory feedback are attempted. Since the project began quite recently, few results has been obtained, so far. This work is a review of main theories on auditory perception and cognition principles, and their merging into an automatic text-to-speech system. This study could also be considered as finalized to the creation of a complete scheme of the human speech/hearing system. Even though this process mainly concerns Speech Synthesis, it is quite sure that its results could be very useful also in Speech Recognition tasks. Knowing how a speaker can change his pronunciation could help the recognizer to set and adapt its models.

References

ANSI S3.5-1997 – “Methods for Calculation of the Speech Intelligibility Index”

Acknowledgment

The research leading to these results has received funding from the European Community’s Seventh Framework Programme under grant agreement n° 213850.11 – SCALE, Speech Communication with Adaptive Learning.