Audio-visual training-aid for speechreading

People with decreasing hearing ability are more dependent on alternative personal communication channels. To 'read and understand' visible articulatory movements of the conversation partner, as done in the process of speechreading, is one possible solution for understanding verbal statements. Training of speechreading skills may be seen as the acquisition of a new, visual language. In this spirit, the presented project demonstrates the conception and implementation of a language laboratory for speechreading that is intended to be employed as an effective audio-visual complement and extension of classroom teaching, but the system may also be used as a new e-learning or, in general, distance learning tool for hearing impaired people. It presents a facial animation on the computer screen with synchronized speech output and is driven by input text sequences in orthographic transcription. The input may later be taken from a speech recognition system. Several basic facial animation systems are available that could be used for this purpose, for example [1], [2], [3]. Our system is based on an existing talking head described in [4] and [5], which is used as the main framework, and on experience on the employment of computer-based communication aids for hearing-impaired, deaf and deaf-blind people [6]. This paper presents the complete system that is composed of a 3D-facial animation with synchronized speech synthesis, a natural language dialogue unit and a student-teacher-training module. Due to the very modular structure of the software package and the centralized event manager, it is possible to add or replace specific modules when needed. The present version of our teacher-student module uses a hierarchically structured composition of important single words and short phrases, supplemented by easy sentences and a guided dialogue. For advanced students, a chatbot system for 'free' dialogue is integrated. The teacher is responsible for designing systematic lessons that resort to formerly learned words or phrases, which is more interesting for the student than a sequential presentation of pre-recorded video material; it also allows the teacher to produce and combine a large number of individual lessons without the need of expensive recording equipment. Our system uses a scene manager to enhance teaching. It allows the creation of different scenarios that are composed of appropriate background images (e.g. kitchen, natural background, traffic corners, railway station, airport), related background sound and lighting conditions. Thereby, it is possible to create an adapted scene (e.g. a restaurant scene with a background image of a restaurant and some kitchen noise) for any specific lesson (e.g. a lesson about the dishes on a menu). The level of difficulty can then be adjusted by changing the loudness of the background noise. The 3D-facial animation is capable of changing appearance and voice in order to match realistic prospects. It is further possible to employ a set of different personified head appearances for the animation, and it is possible to enhance the facial animation by accessories (e.g. a mustache) that might affect the intelligibility of facial and particularly lip movements and, as a consequence, the level of difficulty for the reader. Particularly, also the lighting situation is important for hard-of-hearing students and acoustic reverberation effects of the prospective room for people with low residual hearing. Speechreading requires thorough understanding of spoken language but first and foremost, also of the situational context and the pragmatic meaning of an utterance. To enable the untrained student to have a reasonable context, the head is equipped with a subtitle that makes it possible to show parts of the sentence and to hide specific words with respect to the teacher’s guidelines. It is possible to teach specific words or phrases in the appropriate context without the need of fundamental knowledge of other words. The present version of the training aid can be used for the training of speechreading in English, this as a consequence of the integrated English language models for facial animation and speech synthesis. Nevertheless, the training aid is prepared to handle all possible languages by changing animation and speech synthesis. It has an integrated translator that recognizes the language of the executing computer and loads an optionally available language file. The dialog module is mostly language independent and does not have to be adapted, but, the lessons entered by the teacher must be re-written in the new language. In the paper we present the current version of the training-aid together with results from evaluation experiments with hearing impaired persons and explain functionality and interaction of the modules, the interaction between student, teacher, and virtual teacher, and the structure of the pedagogical approach to teach speechreading with the help of a virtual teacher.