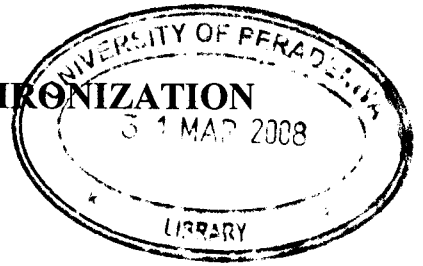


USER DEPENDANT SPEECH BASED LIP SYNCHRONIZATION



A PROJECT REPORT PERSENTED BY

S.M. INOSHA WICKRAMA

✓

to the Board of Study in Statistics and Computer Science of the
POSTGRADUATE INSTITUTE OF SCIENCE

*In partial fulfilment of the requirements
for the award of the degree of*

MASTER OF SCIENCE IN COMPUTER SCIENCE

of the

**UNIVERSITY OF PERADENIYA
SRI LANKA
2007**

614238

USER DEPENDANT SPEECH BASED LIP SYNCHRONIZATION

S. M. I. Wickrama

The University of Peradeniya

Faculty of Engineering

Peradeniya

Sri Lanka

User dependant speech based lip synchronization is an area that is being researched thoroughly. This technology is different from voice puppetry since in voice puppetry the voice puppet learns a facial control model from computer vision of real facial behaviour, automatically incorporating vocal and facial dynamics such as co-articulation, while in lip synchronization the system is based on an Artificial Neural Network and can learn from the user. The more the system is used, the smoother the movements become. The basic focus of this research is to develop a speech based lip synchronization technique which operates in real time. It is anticipated that the final system should be capable of analyzing a speech signal and produce the coordinates of the critical points of the lip.

The work initially started aiming to develop a system that is capable of simulating a 3D face movement with real time user independent speech. However, due to difficulties that arose during the project, which are stated in details in the following chapters, the work was limited to 2D face movement capture and lip synchronization for user dependant speech signals.

The research mainly concentrated on capturing mouth / lip movements and corresponding sounds. It also deals with some attempts at capturing facial movements using limited equipment and some techniques in sound processing which can be used for phoneme recognition.

Several methods were tried out to find out a suitable arrangement for the recording to be made. The marker coordinates on the lip contours are extracted from the recordings and the results are presented in this report.

Then, speech analysis was done using cross-correlation of the phonemes used and it was found that the results were not conclusive enough. As a result, a technique based on Artificial Neural Networks (ANNs) to create a user dependant system was tried. In this technique, the speech analysis consisted of programming into a neural network by feeding direct speech which consisted of 44 phonemes and using the critical points of the lip movements as the desired out come. When the result proved unsatisfactory several other speech processing methods were used to extract the features from the speech signal. These methods included extracting the energy levels of the wave using Discrete Fourier transform and reducing the waves into its wave envelop pattern and results are reported.