Notice of the Final Oral Examination
for the Degree of Master of Applied Science

of

KHALID AMHAD YUSUF AL-HAMMURI

BSc (University of Victoria, 2010)

“Computer Vision-Based Tracking and Feature Extraction
for Lingual Ultrasound”

Department of Electrical and Computer Engineering

Friday, April 26, 2019
11:00 A.M.
Engineering and Computer Science Building
Room 468

Supervisory Committee:
Dr. Poman So, Department of Electrical and Computer Engineering, University of Victoria
(Con-Supervisor)
Dr. Alexandra Branzan Albu, Department of Electrical and Computer Engineering, UVic (Co-
Supervisor)
Dr. Sonya Bird, Department of Linguistics, UVic (Outside Member)

External Examiner:
Dr. Josh Giles, Department of Mechanical Engineering, UVic

Chair of Oral Examination:
Dr. Emmanuel Hérique, Department of French, UVic

Dr. David Capson, Dean, Faculty of Graduate Studies
Abstract

Lingual ultrasound is emerging as an important tool for providing visual feedback to second language learners. In this study, ultrasound videos from five Arabic speakers were collected as they pronounced fourteen Arabic sounds in three different vowel contexts. The sounds were repeated three times to form 630 ultrasound videos. The thesis algorithm was characterized by four steps. First: denoising the ultrasound image by using the combined curvelet transform and shock filter. Second: automatic selection of the tongue contour area. Third: tongue contour approximation and missing data estimation. Fourth: tongue contour transformation from image space to full concatenated signal and features extraction. The automatic tongue tracking results were validated by measuring the mean sum of distances between automatic and manual tongue contour tracking to give an accuracy of 0.9558mm. The validation for the feature extraction showed that the average mean squared error between the extracted tongue signature for different sound repetitions was 0.000858 which means that the algorithm could extract a unique signature for each sound and across different vowel contexts with a high degree of similarity. Unlike other related works, the algorithm showed an efficient and robust approach that could extract the tongue contour and the significant feature for the dynamic tongue movement on the full video frames, not just on the significant single and static video frame as used in the conventional method. The algorithm did not need any training data and had no limitation for the video size or the frame number. The algorithm did not fail during tongue extraction and did not need any manual reinitialization. Even when the ultrasound image recordings missed some tongue contour information, the thesis approach could estimate the missing data with a high degree of accuracy.