



**University
of Victoria**

Graduate Studies

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

of

FENG LIU

BSc (Beijing University of Posts and Telecommunications, 2009)
MSc (Beijing University of Posts and Telecommunications, 2012)

**“Audio Fingerprinting for Speech Reconstruction and Recognition
in Noisy Environments”**

Department of Computer Science

Wednesday, April 12, 2017
9:00 A.M.
Engineering and Computer Science Building
Room 467

Supervisory Committee:

Dr. George Tzanetakis, Department of Computer Science, University of Victoria (Supervisor)
Dr. Kui Wu, Department of Computer Science, UVic (Member)

External Examiner:

Dr. Peter Driessen, Department of Electrical and Computer Engineering, UVic

Chair of Oral Examination:

Dr. Adam Con, School of Music, UVic

Dr. David Capson, Dean, Faculty of Graduate Studies

Abstract

Audio fingerprinting is a highly specific content-based audio retrieval technique. Given a short audio fragment as query, an audio fingerprinting system can identify the particular file that contains the fragment in a library potentially consisting of millions of audio files. In this thesis, we investigate the possibility and feasibility of applying audio fingerprinting to do speech recognition in noisy environments based on speech reconstruction. To reconstruct noisy speech, the speech is divided into small segments of equal length at first. Then, audio fingerprinting is used to find the most similar segment in a large dataset consisting of clean speech files. If the similarity is above a threshold, the noisy segment is replaced with the clean segment. At last, all the segments, after conditional replacement, are concatenated to form the reconstructed speech, which is sent to a traditional speech recognition system.

In the above procedure, a critical step is using audio fingerprinting to find the clean speech segment. To test its performance, we built a landmark-based audio fingerprinting system. Experimental results show that this baseline system performs well in traditional applications, but its accuracy in this new application is not as good as we expected. Next, we propose three strategies to improve the system, resulting in better accuracy than the baseline system. Finally, we integrate the improved audio fingerprinting system into a traditional speech recognition system and evaluate the performance of the whole system.