

## MASTER'S THESIS

### Lip password-based speaker verification system with unknown language alphabet

Zhou, Yichao

*Date of Award:*  
2018

[Link to publication](#)

#### General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

# Abstract

The traditional security systems that verify the identity of users based on password usually face the risk of leaking the password contents. To solve this problem, biometrics such as the face, iris, and fingerprint, begin to be widely used in verifying the identity of people. However, these biometrics cannot be changed if the database is hacked. What's more, verification systems based on the traditional biometrics might be cheated by fake fingerprint or the photo.

Liu and Cheung (Liu and Cheung 2014) have recently initiated the concept of lip password, which is composed of a password embedded in the lip movement and the underlying characteristics of lip motion [26]. Subsequently, a lip password-based system for visual speaker verification has been developed. Such a system is able to detect a target speaker saying the wrong password or an impostor who knows the correct password. That is, only a target user speaking correct password can be accepted by the system. Nevertheless, it recognizes the lip password based on a lip-reading algorithm, which needs to know the language alphabet of the password in advance, which may limit its applications.

To tackle this problem, in this thesis, we study the lip password-based visual speaker verification system with unknown language alphabet. First, we propose a method to verify the lip password based on the key frames of lip movement instead of recognizing the individual password elements, such that the lip password verification process can be made without knowing the password alphabet beforehand. To detect these key frames, we extract the lip contours and detect the interest intervals where the lip contours have significant variations. Moreover, in order to

avoid accurate alignment of feature sequences or detection on mouth status which is computationally expensive, we design a novel overlapping subsequence matching approach to encode the information in lip passwords in the system. This technique works by sampling the feature sequences extracted from lip videos into overlapping subsequences and matching them individually. All the log-likelihood of each subsequence form the final feature of the sequence and are verified by the Euclidean distance to positive sample centers. We evaluate the proposed two methods on a database that contains totally 8 kinds of lip passwords including English digits and Chinese phrases. Experimental results show the superiority of the proposed methods for visual speaker verification.

Next, we propose a novel visual speaker verification approach based on diagonal-like pooling and pyramid structure of lips. We take advantage of the diagonal structure of sparse representation to preserve the temporal order of lip sequences by employ a diagonal-like mask in pooling stage and build a pyramid spatiotemporal features containing the structural characteristic under lip password. This approach eliminates the requirement of segmenting the lip-password into words or visemes. Consequently, the lip password with any language can be used for visual speaker verification. Experiments show the efficacy of the proposed approach compared with the state-of-the-art ones.

Additionally, to further evaluate the system, we also develop a prototype of the lip password-based visual speaker verification. The prototype has a Graphical User Interface (GUI) that make users easy to access.

**Keywords:** visual speaker verification, lip password, language alphabet

# Table of Contents

<b>Declaration</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Table of Contents</b>	<b>v</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>x</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Background & Motivations . . . . .	1
1.1.1 Lip Password-based Visual Speaker Verification . . . . .	1
1.1.2 Lip Password-based Visual Speaker Verification with Hidden Markov Model . . . . .	3
1.1.3 Lip Password-based Visual Speaker Verification with Sparse Coding . . . . .	4
1.2 Thesis Outline . . . . .	6
<b>Chapter 2 Related Works</b>	<b>7</b>
2.1 Visual Feature Extraction . . . . .	7
2.1.1 Static Methods . . . . .	7
2.1.2 Dynamic Methods . . . . .	9

2.1.3	Hybrid Methods . . . . .	9
2.2	Visual Speaker Verification . . . . .	10
2.2.1	Statistical Models . . . . .	10
2.2.2	Sparse Coding . . . . .	10
2.3	Lip password-based Speaker Verification . . . . .	11

**Chapter 3 Interest Interval Model and Overlapping Subsequence Matching Approach to Lip Password-based Visual Speaker Verification with Hidden Markov Models** **12**

3.1	Introduction . . . . .	12
3.2	Overview of Hidden Markov Model . . . . .	14
3.3	Interest Interval Model with HMM . . . . .	15
3.3.1	Feature Extraction . . . . .	16
3.3.2	Interest Intervals . . . . .	17
3.3.3	Interest Interval Model . . . . .	19
3.3.4	Experiments . . . . .	22
3.4	Overlapping Subsequence Matching with HMM . . . . .	29
3.4.1	Feature Extraction . . . . .	29
3.4.2	Overlapping Subsequence Matching . . . . .	30
3.4.3	Experiments . . . . .	33
3.5	Summary . . . . .	38

**Chapter 4 Diagonal-like Pooling and Pyramid Structure of Lips Based on Sparse Coding** **40**

4.1	Introduction . . . . .	40
4.2	Overview of Sparse Coding Learning Framework . . . . .	41
4.3	The Proposed Method . . . . .	42
4.3.1	Diagonal-like Pooling . . . . .	43
4.3.2	The Pyramid Structure of Lips . . . . .	45
4.4	Experiments . . . . .	47

4.4.1	Database . . . . .	47
4.4.2	Experiment Protocol . . . . .	47
4.4.3	Experiment Results . . . . .	49
4.4.4	Performance Comparison with the State-of-the-art . . . . .	51
4.5	Conclusion . . . . .	52
<b>Chapter 5 The lip-password based visual speaker verification proto-</b>		
	<b>type</b>	<b>53</b>
5.1	Introduction . . . . .	53
5.2	System Framework & Components . . . . .	54
5.2.1	System Framework . . . . .	54
5.2.2	Graphical User Interface . . . . .	55
5.2.3	User Database . . . . .	58
5.3	System Implementation . . . . .	58
5.3.1	Localization . . . . .	59
5.3.2	Feature Extraction . . . . .	59
5.3.3	Verification Algorithms . . . . .	60
5.4	Summary . . . . .	60
<b>Chapter 6 Conclusion and Future Work</b>		<b>61</b>
6.1	Conclusion . . . . .	61
6.2	Future Work . . . . .	63
<b>Bibliography</b>		<b>64</b>
<b>Curriculum Vitae</b>		<b>71</b>