# Lingyun Yu

## PERSONAL STATEMENT

My research interests span Talking face generation, Articulatory movements-driven 3D Talking Head, Human-Computer Interaction and Video synthesis. The noteworthy research project of mine is to generate realistic talking heads synchronized with the given audio or text input. I am skilled at python and matlab coding, and experienced of Linux system.

## CONTACT INFORMATION


Phone: +86-13966763354 Email: yuly@mail.ustc.edu.cn

Homepage: http://home.ustc.edu.cn/~yuly WeChat: 13966763354


## EDUCATION

**09/2015-Present** **Doctor of Control Science and Engineering**
University of Science and Technology of China (USTC)

**09/2011-06/2015** **Bachelor of Electrical Engineering and Automation**
China University of Mining and Technology (CUMT)

## RESEARCH EXPERIENCES

**10/2017-Present** ***Talking Face Generation from Text or Audio input*** *Independent Completor*

- To overcome this problem, we decompose the talking face generation task into two steps: mouth landmarks prediction and video synthesis.
- **The time-delayed LSTM** is adopted to predict accurate mouth landmarks.
- A network named Face2Vid is proposed to generate video frames conditioned on the predicted mouth landmarks. In Face2Vid, **the optical flow** is employed to model the temporal dependency between frames, meanwhile, **a self-attention mechanism** is introduced to model the spatial dependency across image regions.

*One Journal paper accepted by **ICDM 2019*** (Regular paper, 9.08% acceptance rate).
*One conference paper submitted to **IEEE TCSVT**.*

**05/2016-09/2017** ***Articulatory Synchronicity from Text and Audio Inputs*** *Independent Completor*

- We propose a new network architecture for articulatory movement prediction with both text and audio inputs, called **BLTRCNN.**
- The bottleneck network is adopted to extract the compact bottleneck features as the complementary linguistic features for better performance.
- **C**ombining **CNN, LSTM and skip connection** can not only acquire local higher-level features but also learn long-range of context information.
- **Combining acoustic features and linguistic features** as inputs is able to contain more information to boost the performance.

*One Journal paper accepted by **IEEE TMM***
*Conference papers accepted by **MMM2019*** and ***VCIP2018**.*

**09/2017-11/2017** ***The personalized co-articulation rule statistics*** *Independent Completor*

- We aim at studying the personalized co-articulation rule among neighboring phonemes by the representation of the mouth shape with both text and audio information as inputs.
- **A time-delayed LSTM** is used to model the mapping from the linguistic and acoustic features to the mouth landmarks.
- **By statistical knowledge**, we calculate and study the co-articulation rule among neighboring

phonemes.

**02/2015-02/2016** ***Articulatory movements-driven 3D pronunciation system.*** *Independent Completor*

- We proposes an emotional text-driven 3D visual pronunciation system for Mandarin Chinese by generating the emotional articulatory movement trajectory.
- The articulatory movements are predicted by HMM.
- We analyze and summarize the variation of the articulatory movements under different emotions.

*One Journal paper accepted by **Multimedia Tools and Applications**.*

*Conference papers accepted by **CCPR2016.***

**07/2019-04/2019** ***The "AI Meets Beauty" Challenge in ACM MM2019*** *Project Participant*

- We propose a novel Generalized-attention Regional Maximal Activation of Convolutions (GR-MAC) descriptor to boosts retrieval performance.
- **Attention mechanism** is introduced to assign larger weights for target regions.
- Awarded **the 1st place** in the Grand Challenge of AI Meets Beauty.

*Conference papers accepted by **ACM MM 2019**.*

## INTERNSHIP EXPERIENCES

**07/2017-01/2019** ***Iflytek Co.,Ltd***

- Given an arbitrary speech clip or text information as input, we aim to generate a talking face video with accurate lip synchronization.

## PUBLICATIONS

- **Lingyun Yu** et al. Mining Audio, Text and Visual information for Talking face generation **IEEE ICDM 2019 (Regular paper, 9.08% acceptance rate)**
- **Lingyun Yu**, Jun Yu, and Qiang Ling. 2018. BLTRCNN Based 3D Articulatory Movement Prediction: Learning Articulatory Synchronicity From Both Text and Audio Inputs. **IEEE Transactions on Multimedia（IF:5.452）.**
- **Lingyun Yu** et al. Multimodal Inputs Driven Talking Face Generation With Spatial-Temporal Dependency (**IEEE Transactions on Circuits and Systems for Video Technology** ) **(major revision IF: 4.046)**
- **Lingyun Yu**, Yu J, Wang Z. A realistic 3D articulatory animation system for emotional visual pronunciation[J]. **Multimedia Tools and Applications**, 2017, 76(18): 19241-19262 **(IF:2.101).**
- **Lingyun Yu**, Yu J, Ling Q. Deep Neural Network Based 3D Articulatory Movement Prediction Using Both Text and Audio Inputs[C]//International Conference on Multimedia Modeling. Springer, Cham, 2019: 68-79.
- **Lingyun Yu**, Yu J, Ling Q Synthesizing 3D Acoustic-Articulatory Mapping Trajectories: Predicting Articulatory Movements by Long-Term Recurrent Convolutional Neural Network. IEEE VICP (2018)
- **Lingyun Yu**, Changwei Luo, and Jun Yu. "An Emotional Text-Driven 3D Visual Pronunciation System for Mandarin Chinese." Chinese Conference on Pattern Recognition. Springer, Singapore, 2016.
- Yu J, **Lingyun Yu**. Synthesizing Photo-Realistic 3D Talking Head: Learning Lip Synchronicity and Emotion from Audio and Video[C]//2018 25th IEEE International Conference on Image Processing. IEEE, 2018: 1448-1452.
- Jun Yu et al., **Lingyun Yu,** Beauty Pro duct Retrieval Based on Regional Maximum Activation of Convolutions with Generalized Attention. ***ACM MM 2019*** (Accept)
- Yu J, Luo C, **Lingyun Yu**, et al. Facial video coding/decoding at ultra-low bit-rate: a 2D/3D model-based approach[J]. **Multimedia Tools and Applications, 2016: 1-21. (IF: 2.101)**
- Luo C, Li R, **Lingyun Yu**, et al. Automatic Tongue Tracking in X-Ray Images[J]. Chinese Journal of Electronics, 2015, 24(4): 767-771. **(IF: 0.945)**
- Chuanbin Liu, Hongtao Xie, Zhengjun Zha, **Lingyun Yu**, Zhineng Chen, and Yongdong Zhang. Bidirectional Attention-Recognition Model for Fine-grained Object Classification. **IEEE Transactions on Multimedia.**

**(Accepted IF：5.452 SCI II)**

## FUNDS

**National Natural Science Foundation of China (U1736123)**

**National Natural Science Foundation of China (61572450)**

**National Natural Science Foundation of China (61303150)**

**Huawei Innovation Research Program - Flagship Program (ES2100100098)**

**Iflytek Co.,Ltd. Technology research and development project (ES2100100099)**

## HONORS & AWARDS

**2019/07** **1st place, ACM Multimedia “AI Meets Beauty” challenge**

**2017/09** **National Scholarships, USTC**

**2017/11** **Travel Award, 2018 IEEE VCIP**

**2015/05** **Excellent graduates, CUMT**

**2014/10** **Exam-Free Postgraduate to USTC**