**Иванько Денис Автоматическое распознавание аудиовизуальной русской речи**

ОГЛАВЛЕНИЕ ДИССЕРТАЦИИ

кандидат наук Иванько Денис

СОДЕРЖАНИЕ

Сокращения

Реферат

1. Introduction

1.1 Automatic speech recognition

1.2 Motivation

1.3 Thesis contribution

1.4 Outline

2. Backgrounds and related research

2.1 Automatic acoustic speech recognition

2.1.1 Acoustic features extraction

2.1.2 Acoustic modeling

2.1.3 Language modeling

2.2 Automatic visual speech recognition

2.2.1 Region-of-interest detection

2.2.2 Visual features extraction

2.2.3 Visual speech recognition

2.3 Multimodal speech recognition

2.3.1 Audio-visual features extraction

2.3.2 Audio-visual fusion approaches

2.3.3 Audio-visual fusion techniques

2.4 Proposed approach to practical lip-reading system implementation

2.5 Summary

3. General methodology and contribution to the state-of-the-art

3.1 Acoustic speech processing

3.2 Visual speech processing

3.2.1 Haar classifiers-based method for region-of-interest detection

3.2.2 Proposed modification of a method for lip region detection

3.2.3 active appearance model-based method for region of interest detection116

3.2.4 Pixel-based visual features extraction

3.2.5 Proposed geometry-based vsual features extraction method

3.3 Modalities fusion and modeling

6

3.3.1 Hidden Markov models and Gaussian mixtures models

3.3.2 Coupled hidden Markov models

3.3.3 Hybrid approach to speech recognition

3.3.4 End-to-end approach

3.4 Decoding and evaluation

3.4.1 Decoding

3.4.2 Evaluation metrics

3.5 Summary

4. Data collection and tools analysis

4.1 Data

4.1.1 Audio-visual and visual-only speech datasets

4.1.2 Distinctive features of the Russian audio-visual speech

4.1.3 Software-hardware complex for database recording

4.1.4 HAVRUS corpus description

4.1.5 GRID dataset

4.2 Tools

4.2.1 Toolkits

4.2.2 Deep learning frameworks

4.2.3 Computer vision libraries

4.3 Summary

5. Experimental setups and evaluations

5.1 Experimental setup

5.1.1 Building traditional audio-visual speech recognition system

5.1.2 Building hybrid audio-visual speech recogniton system

5.1.3 Building End-to-end visual speech recognition system

5.2 Evaluation experiments

5.2.1 Experiments with the frame rate

5.2.2 Experiments in acoustically noisy environments

5.2.3 Experiments with viseme classes

5.2.4 Experiments with visual features

5.2.5 Experiments with different architectures of speech recognition systems186

5.3 Summary

6. Conclusion and future directions

6.1 Overall summary

7

6.2 Thesis contributions

6.2.1 Theoretical

6.2.2 Practical

6.2.3 Experimental

6.3 Future directions

Appendix

References

Приложение А. Тексты публикаций