**Мамырбаев Оркен1\*** [morkenj@mail.ru](mailto:morkenj@mail.ru)

**Калимолдаев Максат1** [mnk@ipic.kz](mailto:mnk@ipic.kz)

**Амиргалиев Едилхан1** amir\_ed@mail.ru

**Мусабаев Рустам1** rmusab@gmail.com

**Alimhan Keylan2** keylan@live.jp

1Алматы, Институт информационных и вычислительных технологий КН МОН РК

2Tokyo, Tokyo Denky University

**Метод разработки системы многомодального распознавания казахской речи**

В настоящее время за рубежом многомодальные методы распознавания речи уже используются в некоторых прикладных областях: картографических системах, системах виртуальной реальности, медицинских системах, робототехнике, web-приложениях, и т.д. [1]. Помимо этого, многомодальные методы распознавания речи могут быть полезны в мобильных устройствах, где имея ограничения для ввода информации с помощью обычной клавиатуры.

Исследования, посвященные распознаванию речи, лица, положения человека в окружающем пространстве ведутся уже более полувека. Однако, системы объединяющие различные способы ввода информации в единой форме стали разрабатываться совсем недавно. Такие распознающие системы используют многомодальных (мультимодальных) методов распознавания речи. Многомодальные методы распознавания речи обрабатывают данные полученных из двух и более каналов ввода информации например, как речь и движения губы [2].

Созданная нами система рассматривает двухмодальный (речь и губы) метод распознавания. По первому каналу поступает речевой сигнал от микрофона, а из второго канала поступает сигнал от видео камеры Кinect, описывающий движение губы. Разработанная многомодальная система, получая информацию из двух каналов совместно обрабатывая их, позволяет более качественно распознавать речь. В результате, установлено, что качество распознавания речи на основе двухмодального подхода выше, чем качество, полученное от отдельно взятой речи, учитывающая в системах [3, 4].

В данной работе исследуется метод двухмодального распознавания речи, основанный на анализе речевого сигнала и изображения лица говорящего. Интеграция данных подходов рассматривается с точки зрения комбинирования классификаторов, когда каждый метод обучается отдельно по своей модальности. Распознавание по речевому сигналу обладает относительно высоким, но недостаточным уровнем точности распознавания. По результатам теоретических оценок и проведенных экспериментов установлено, что распознавании речи в комбинации с менее точным методом распознавания, как распознавание по изображению лица говорящего, качество распознавания может быть улучшено.

1. Jain A.K., Ross A., Prabhakar S. An introduction to biometric recognition // IEEE Trans. Circuits Syst. Video Technol. Vol. 14, 2004, P. 4–20.
2. Raghavendra R. Robust Algorithms for Person Verification using Multimodal Biometrics // Ph.D Thesis, Feb.
3. Kalimoldayev M. N., Mamyrbayev O. J., Mussabayev R. R., Orazbekov J. N. Segmentation and speech signal processing algorithm using the average frequency level crossing // Problems of Informatics, 1(22), 2014, P. 73-82.
4. Cootes T., Edwards G., Taylor C. Active appearance models // In Proceedings of the European Conference on Computer Vision, Vol. 2, 1998, P. 484–498.