

УДК 57. 08

Биотехническая система распознавания фонем русской речи по изображению губ

07, июль 2012

Самойлов Д.С.

*Студент,
кафедра «Биомедицинские технические системы»*

*Научный руководитель: Аполлонова И.А.,
к. т. н., доцент кафедры «Биомедицинские технические системы»*

МГТУ им. Н.Э. Баумана
Samoylov.work@gmail.com

1.1 Комплекс параметров артикуляции на основе контура внутренней области губ.

Виземы (визема - визуальная реализация фонемы) можно описать совокупностью параметров артикуляции – геометрическими признаками губ (высота, ширина, площадь, периметр контура и т.д.).

Для автоматизации процесса вычисления комплекса параметров артикуляции выбран метод аппроксимации автоматически выделенной внутренней области губ эллипсом.

Для автоматического выделения параметров артикуляции выбран внутренний контур губ и следующие параметры:

1. ширина внутреннего контура;
2. высота внутреннего контура.

1.2 Алгоритм выделения внутренней области губ и аппроксимации области эллипсом

Алгоритм выделения внутренней области губ основан на анализе цветового пространства. Первоначально выделяется область интересов, в которой располагаются губы человека (рис.1).



Рис.1. Автоматически выделенная область губ

Затем происходит преобразование из цветового пространства RGB в цветовую модель HSV (рис.2).

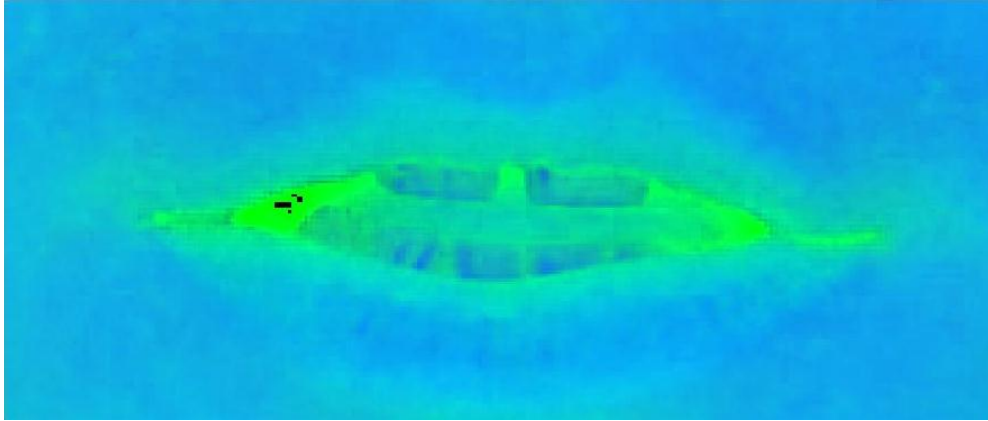


Рис.2. Изображение области губ в цветовой модели HSV

HSV (англ. Hue, Saturation, Value — тон, насыщенность, значение) — цветовая модель, в которой координатами цвета являются:

- Hue — цветовой тон, (например, красный, зелёный или сине-голубой). Варьируется в пределах $0 - 360^\circ$, однако иногда приводится к диапазону $0 - 100$ или $0 - 1$.
- Saturation — насыщенность. Варьируется в пределах $0 - 100$ или $0 - 1$. Чем больше этот параметр, тем «чище» цвет, поэтому этот параметр иногда называют чистотой цвета. А чем ближе этот параметр к нулю, тем ближе цвет к нейтральному серому.
- Value (значение цвета). Также задаётся в пределах $0 - 100$ и $0 - 1$.

Разность цветовых компонент S и V позволяет получить изображение, подходящее для выделения внутренней области губ (рис.3).

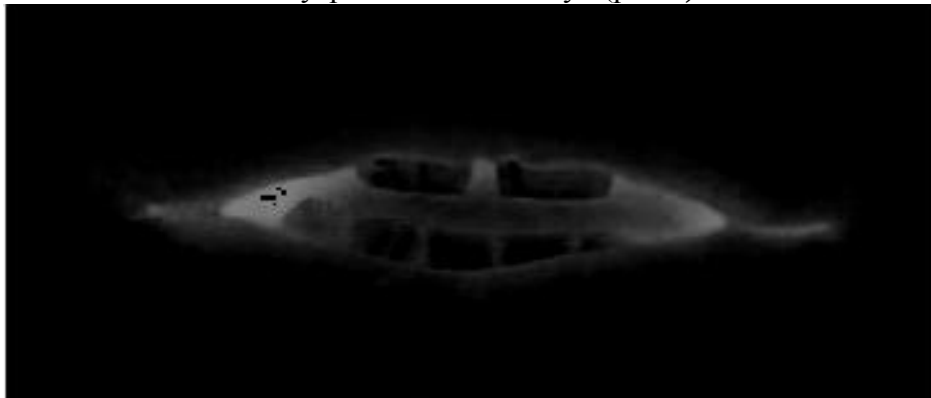


Рис.3. Изображение разности цветовых компонент S и V

После, происходит бинаризация полученной разности (рис.4).



Рис.4. Бинаризация изображения

Выделенный контур аппроксимируется эллипсом при помощи метода наименьших квадратов (рис.5). Метод наименьших квадратов — метод нахождения оптимальных параметров линейной регрессии, таких, что сумма квадратов ошибок (регрессионных остатков) минимальна.

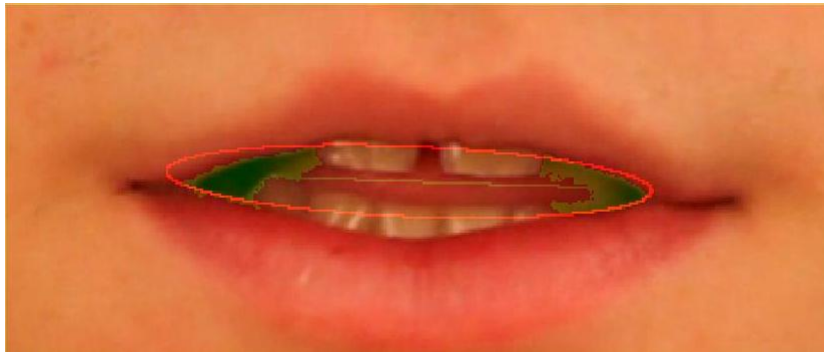


Рис.5. Построение эллипса, аппроксимирующего внутренний контур губ

Полуоси вычисленного эллипса являются параметрами артикуляции.

1.3 Описание исследования распознавания фонем

Исследование распознавания фонем началось с поиска фонетически сбалансированного текста, наиболее подходящего для оценки возможности обнаружения фонем по выбранным нами параметрам артикуляции.

Пример фонетически сбалансированных предложений:

Вместе с зимой и холодами к селянам пришли всякие беды.

Хиппи обычно называют людей, бросающих обществу вызов своей пассивностью и бездеятельностью.

Продавец должен учитывать не только экономические, но и психологические факторы цены.

Снято видео с чтением этого текста для дальнейшей обработки.

В последующей обработке полученных данных, после применения вышеописанного алгоритма к снятому видео, найдена зависимость для определенных групп фонем (рис.6).

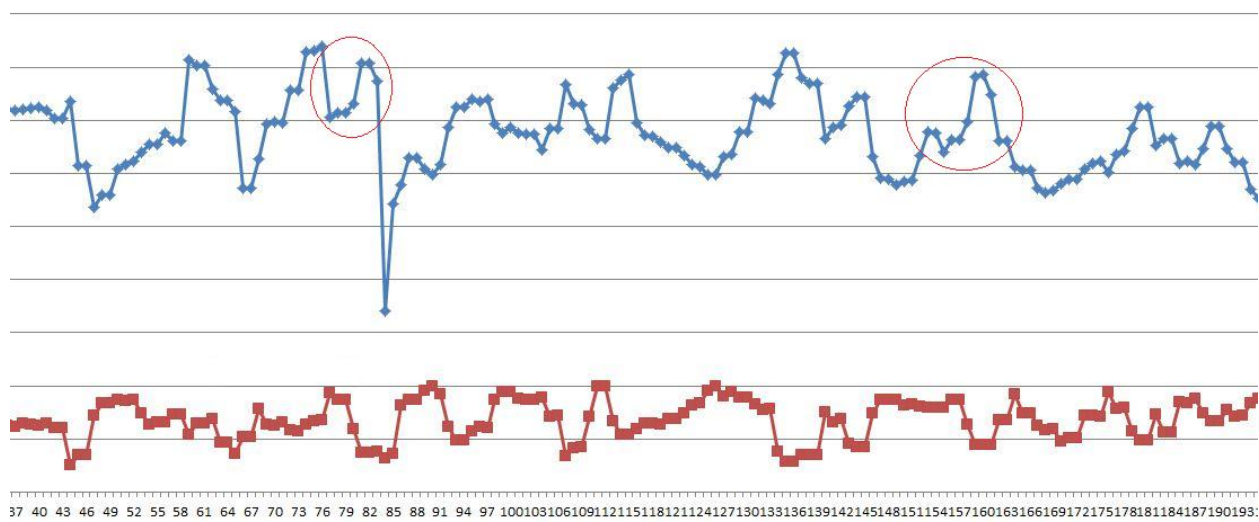


Рис.6. График с зависимостями для фонемы «в»

Для дальнейшего исследования составлен текст, в который входят слова, сочетающие в себе «губные» согласные фонемы и сочетания их с различными гласными фонемами.

Пример текста

бе(пе) еб(еп) бедро, зебра, персик, лепить

бо(по) об(оп) боль, польза, зоб, топот

Снята база данных с чтением этого текста. После анализа вновь полученных данных зависимость подтвердилась (рис.7), и решено разработать алгоритм для автоматического определения необходимых фонем в словах.

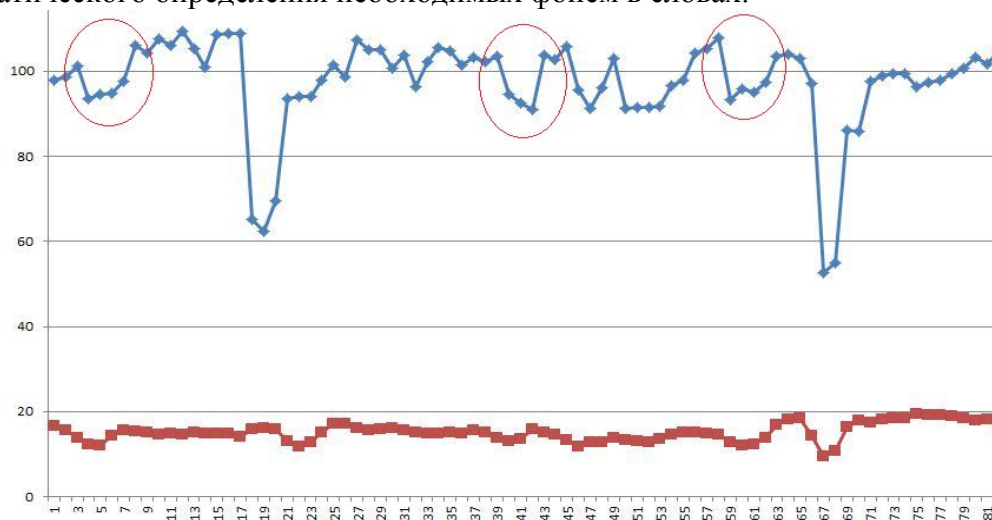


Рис.7. График с подтвержденной зависимостью для фонемы «в»

Вновь снята база данных с чтением этого текста несколькими людьми. И каждое сочетание фонем произнесено отдельно каждым участником по 3 раза для получения «эталона» (рис.8).

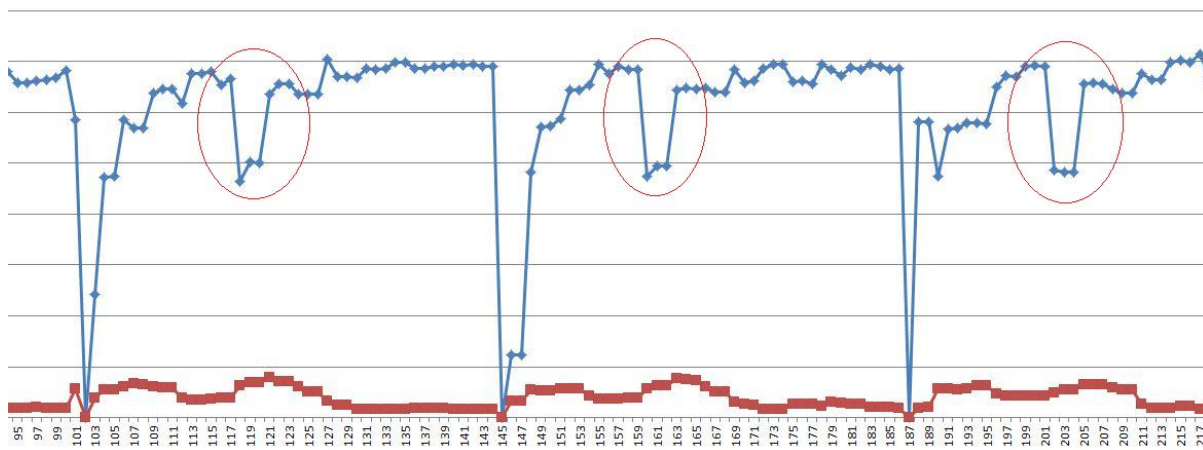


Рис.8. График «эталона» для сочетания фонем «в» и «е»

Обработанная база данных дала нам «эталон» по параметрам артикуляции на каждую из визем, что позволило провести исследование по их корреляции. Исследование заключалось в поиске отклика одной виземы на все остальные. Результаты исследования были занесены в таблицу, после которой были сделаны выводы.

Проведенные исследования по корреляции опорных визем дали следующие результаты:

- 1) 85 % исследуемых визем обнаруживаются по корреляции с самими с собой.

Данный результат можно улучшить путем увеличения количества рассматриваемых параметров артикуляции, а так же улучшения алгоритма выделения внутренней области губ и ее аппроксимации.

- 2) 15% исследуемых визем дают ложные отклики по корреляции с другими виземами в пределах 30-50%.

После проведенного анализа этого результата можно сказать, что ложные отклики получены потому что эти виземы визуально очень схожи и соответственно полученные параметры артикуляции имеют близкие значения.

Список используемой литературы

1. Миронова Э.В. Оценка навыка чтения с губ. М., 1980.
2. Р. Гонсалес, Р. Вудс. Цифровая обработка изображений. М., 2005.
3. Мурыгин К. В. Концепция системы распознавания речи на основе чтения по губам. «Искусственный интеллект» - 2009 –№ 2. – С. 116-123.