

## РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ ГЛАСНЫХ ФОНЕМ И СОСТАВНЫХ ЗВУКОВ С ЦЕЛЮ БЕЗДИКТОРНОГО РАСПОЗНАВАНИЯ РЕЧИ

С целью решения проблемы дикторнезависимого распознавания речи на фонемном уровне, при создании алгоритма, описывающего речевой сигнал как детерминированную модель, а не вероятностную (скрытые марковские процессы и нейросети) встает вопрос о необходимости глубокого исследования гласных фонем и составных гласных звуков. Решению данной проблемы и посвящена данная статья.

Согласно алгоритму распознавания гласных фонем и составных гласных звуков русского языка как отдельно произносимых, так и в контексте, была разработана программа на языке программирования Pascal, позволяющая обрабатывать файлы по пути: первичная обработка речевого сигнала с целью выделения частоты основного тона, по которой идет формирование низкочастотного сигнала [1, 2]; автоматическое выделение сегментов (периодов анализа) [3, 4, 5] и сегментный скользящий спектральный анализ всего сигнала с определением периодов подобия [7]; построение кодовой комбинации (частотного портрета) по периодам подобия [6, 7]; сравнение кодовой комбинации со шкалой анализа с целью принятия решения о принадлежности фонемы [8].

Для удобной и наглядной работы программы предусматривается прорисовка на экране исходного сигнала и сигналов, полученных в ходе математической обработки. Встроенный математический аппарат позволяет сравнивать группы сигналов друг с другом последовательно, производя: вычитание сигналов, интегрирование, дифференцирование и фильтрацию сигналов, определение энергии сигнала, аппроксимацию сигналов (полиномы и сплайны), что дает возможность проводить статистический анализ сигналов по всей создаваемой

Речевой сигнал гласных фонем

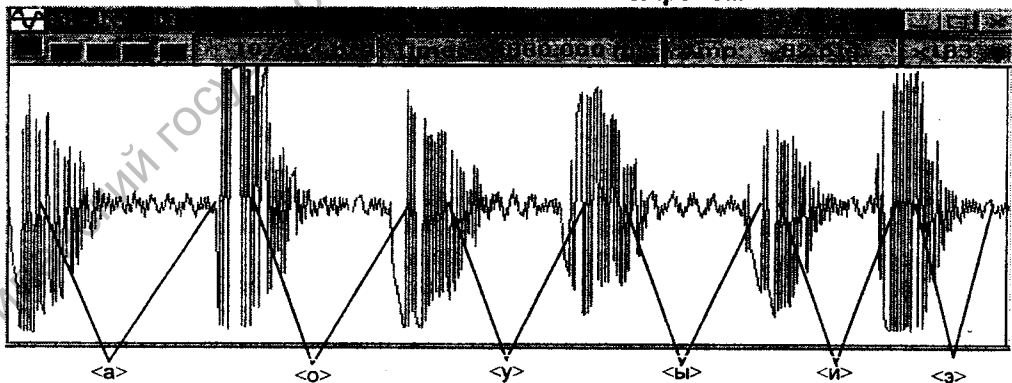


Рис. 1.

## Речевой сигнал составных гласных звуков, обозначаемых буквами

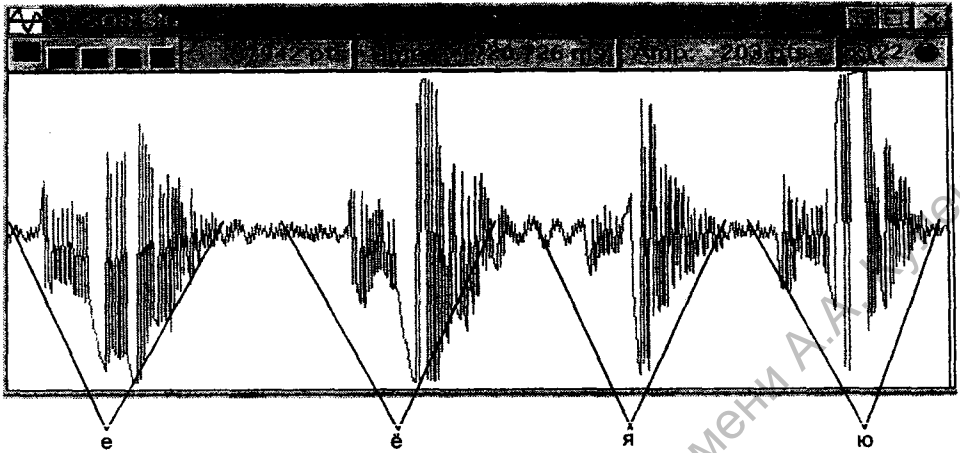


Рис. 2.

базе данных для каждой группы дикторов (мужчин, женщин, детей) для выявления общих и отличительных характеристик.

На рис. 1 представлены речевые сигналы гласных фонем, к которым относятся <a>, <o>, <y>, <ы>, <и>, <э>. Особое внимание было уделено исследованию составных гласных звуков, обозначаемых буквами: я, ё, е, ю. На рис. 2 представлены речевые сигналы этих звуков.

По данным рисункам уже визуально можно судить о различиях в реализации этих фонем. При исследовании речевых сигналов, содержащих данные составные гласные звуки, было обнаружено, что эти звуки имеют несколько ярко выраженных периодов. Причем каждый из периодов отличается не только на слух, но и визуально (см. рис. 1-2).

Данные гласные изучались как самостоятельно, так и в контексте согласный-гласный с использованием созданных базы данных и программного обеспечения.

В результате исследований было установлено, что:

1. звуковое значение буквы я состоит из четырех звуков: [ы], [й], [э], [а] (табл. 4.1);
2. звуковое значение буквы ё состоит из пяти звуков: [и], [й], [у], [о], [а] (табл. 4.2);
3. звуковое значение буквы е состоит из четырех звуков: [и], [й], [и], [э] (табл. 4.3);
4. звуковое значение буквы ю состоит из пяти звуков: [и], [ы]→[и], [й], [у] (табл. 4.4).

В таблицах 1-4 представлены результаты исследования составных гласных звуков для диктора Ж1 (здесь и далее Ж означает, что диктором является женщина; М - что диктором является мужчина).

Как видно из рис. 2, составные гласные звуки имеют четко выраженные границы составляющих периодов. После математической обработки эти периоды анализируются в соответствии с предлагаемым алгоритмом. После проведения скользящего спектрального анализа снимались показания для четырех локальных по амплитуде максимумов (в таблице они имеют порядковые номера 1, 2, 3, 4).

Таблица 3

## Звуковое значение буквы е. Диктор Ж1

Слыш. звук	1		2		3	
	оберт.	част.	оберт.	част.	оберт.	част.
и	1 ⊕	185,294	13	2408,824	19	3520,588
	1 ⊕	170,930	22	3760,465	32	5469,767
й	2 ⊕	386,842	13	2514,47	18	3481,579
	2 ⊕	397,297	12	2383,784	17	3377,027
и	2 ⊕	400,909	12	2405,455	14 !	2806,364
	2 ⊕	408,333	12	2450	14	2858,333
э	2 ⊕	432,353	10	2161,765	13	2810,294
	2 ⊕	450	9	2025	12	2700

Таблица 4

## Звуковое значение буквы ю. Диктор Ж1

Слыш. звук	1		2		3	
	оберт.	част.	оберт.	част.	оберт.	част.
и	2 ⊕	390,265	5	975,664	15	2926,991
ы-и	1 ⊕	185,294	6	1111,76	16	2964,706
й	1 ⊕	183,750	3	551,250	17	3123,750
й	1 ⊕	202,294	11	2225,229	15	3034,404
у	1 ⊕	237,097	8	1896,774	–	–
	1 ⊕	227,320	6	1363,918	9	2045,876
	1 ⊕	232,105	5	1160,526	8	1856,842
	1 ⊕	225	4	900	10	2250
	1 ⊕	218,317	–	–	–	–
	1 ⊕	227,320	11 !	2500,515	–	–

Для каждого максимума определяется частота и его порядковый номер в спектре (в таблице – обертон), где значком  $\oplus$  отмечается глобальный по амплитуде для данного спектра максимум.

Таблица 1

**Звуковое значение буквы я. Диктор Ж1**

Слыш. звук	1		2		3	
	оберт.	част.	оберт.	част.	оберт.	част.
ы	2 $\oplus$	376,923	6	1130,769	8	1507,692
й	1 $\oplus$	182,231	4	728,926	6	1093,388
	1 $\oplus$	253,448	14	3548,276	24	6082,759
э	2 $\oplus$	400,909	11	2205	16	3207,273
	3 $\oplus$	601,364	10	2004,545	14	2806,364
а	2	370,588	4 $\oplus$	741,176	10	1852,941
	2	352,8	4 $\oplus$	705,6	6	1058,4
	2	347,244	4 $\oplus$	694,488	9	1562,598
	1	173,622	4 $\oplus$	694,488	9	1562,598

Таблица 2

**Звуковое значение буквы ё. Диктор Ж1**

Слыш. звук	1		2		3	
	оберт.	част.	оберт.	част.	оберт.	част.
и	2 $\oplus$	376,923	12	2261,538	15 !	2826,923
й	2 $\oplus$	339,231	14	2374,615	17	2883,462
у	2 $\oplus$	412,150	9	1854,673	12	2472,897
	2 $\oplus$	420	7	1470	10	2100
	2 $\oplus$	424,038	5	1060,096	9	1908,173
о	2 $\oplus$	436,634	12 !	2619,802	–	–
	3 $\oplus$	661,5	13 !	2866,5	–	–
а	2 $\oplus$	450	5	1125	–	–

После сопоставления полученной частоты обертона со шкалой анализа [8] принимается решение о принадлежности, в результате чего и получили те гласные звуки, которые определяют составной гласный звук, обозначенный буквой я, т.е. звуки: ы, (с переходом на [и]), й, э, а. В таблицах 1–4 знаком "!" отмечается равный (или стремящийся к нему) по амплитуде обертон с глобальным максимумом.

Аналогичным образом предлагается распознавать согласные фонемы, в частности, исследовать сонанты. Отличие их спектральной характеристики от гласных фонем характеризуется порядковым номером второго и третьего обертона. Так, например, положение обертонов для фонемы <а>: 2-5-10 (11, 12), а для <н>: 1-8 (9)-13 (14, 15, 16) или 1-7-9 (10, 14). Если исходить из формулы

$$\omega_i = i \cdot \omega_1,$$

где  $\omega_1$  – частота (Гц) основного тона;  $\omega_i$  – частота  $i$ -ой гармоники;  $i$  – порядковый номер гармоники, то для сонант можно воспользоваться той же шкалой анализа [8], что и для гласных звуков.

Особое внимание уделялось альвеолярной щелевой согласной <с>, которую можно определить по численному значению частоты и расположению обертонов в спектре. Всего рассматривалось 15 обертонов  $U^P(\omega)$ , для которых отмечалось наличие локальных и глобальных максимумов:

$$U^P(\omega) = [\omega_1^{\max}, \omega_2^{\max}, \dots, \omega_i^{\max}] \rightarrow \max \text{ при } i=1 \div 15.$$

Установлено, что глобальный  $\omega^{\max}$  максимум для <с> лежит в интервале (Гц):  $2500 \leq \omega^{\max} \leq 6500$ .

Таким образом, по разработанному алгоритму для распознавания речевых звуков русского языка были исследованы:

- 1) гласные фонемы <а>, <о>, <у>, <и>, <ы>, <э>;
- 2) составные гласные звуки, обозначаемые буквами я, ё, е, ю;
- 3) согласная сонанта <н>;
- 4) альвеолярная щелевая согласная <с>.

Установлено, что для каждой гласной фонемы существует свой частотный диапазон ее реализации, который полностью раскладывается по используемой в программе шкале анализа [8]. Данная шкала является внешней по отношению к сигналу и носит экспоненциальный характер.

В ходе проведенного анализа результатов было установлено, что для распознавания гласных и согласных фонем по способу и месту образования можно применять разработанное программное обеспечение (ПО) для фонемного автоматического распознавателя дикторнезависимого непрерывного речевого сигнала. Достоинством ПО и изложенной методики распознавания является отсутствие эталонов анализа, сигнал на каждом этапе сравнивается сам с собой по энергии с целью выявления закономерностей его изменения [6, 7, 8].

Полученные результаты могут быть использованы для дикторнезависимого распознавания гласных и согласных фонем и звуков в непрерывном речевом сигнале русского и белорусского языков.

#### ЛИТЕРАТУРА

1. Иванова И.Д., Цупрев Н.И., Шлапаков С.А. Дифференцирование – как способ проявления информации в речевом сигнале // Вестник ВГУ. – 1998. – № 2 (8), – С. 70–74.

2. **Иванова И.Д.** Решение задачи унификации речевого сигнала при помощи дифференцирования // *Материалы междунар. конф. "Еругинские чтения V": В 2-х ч.* – Могилев, 1998. – Ч. 2. – С. 103. – Рус.
3. **Иванова И.Д.** Способ определения интервала для скользящего спектрального анализа / *МогТИ.* – Могилев, 1997. – 8 с.: ил. – Библиогр. 3 назв. – Рус. – Деп. в БелИСА 31.12.97, № Д199751. Минск: Реф. сб. неопубл. работ вып. № 8.
4. **Иванова И.Д.** Алгоритм определения ядра и периодов подобия в речевом сигнале / *МогТИ.* – Могилев, 1997. – 9 с.: ил. – Библиогр. 5 назв. – Рус. – Деп. в БелИСА 12.12.97, № 199746. Минск: Реф. сб. неопубл. работ вып. № 8.
5. **Иванова И.Д.** Автоматическая сегментация речевого сигнала с целью получения частотных признаков гласных фонем // *Доклады 1 Междунар. конф. "Цифровая обработка сигналов и ее применение". DSPA'98: В 4-х т.* – М., 1998. – Т. 4. – С. 42–48. – Рус.
6. **Иванова И.Д.** Исследование фонем по способу и месту образования // *Техника и технология пищевых производств: Междунар. науч.-техн. конф.: Тез. док.* – Могилев, 1998. – С. 293. – Рус.
7. **Иванова И.Д., Цупрев Н.И.** Способ анализа и система кодирования сегментов речевых сигналов // *Техника и технология пищевых производств: Междунар. науч.-техн. конф.: Тез. док.* – Могилев, 1998. – С. 305. – Рус.
8. **Иванова И.Д.** Выбор и обоснование метода идентификации гласных фонем в речевом сигнале // *Материалы междунар. конф. по статистической динамике информационных систем и потоков.* 22-23 мая 1998. – Бобруйск (Беларусь), 1998. – С. 26–30. – Рус.