

УДК 004.934.1'1

Е.Е. Федоров, И.А. Шевцова

Донецкий государственный институт искусственного интеллекта, Украина

Численное исследование шипящих согласных звуков

При создании систем распознавания речи важную роль играет выбор признаков, на основе которых формируются эталоны звуков речи. Для решения этой задачи в статье приведено численное исследование шипящих согласных звуков.

Постановка проблемы

В современной отечественной и мировой практике активно ведутся разработки естественно-языковых систем общения «человек–компьютер», одной из составных частей которых может быть система распознавания речи. При создании системы распознавания необходимо решить задачу выбора словаря признаков. Для решения этой проблемы авторами проведено численное исследование шипящих звуков речи русского языка, результаты которого представлены в данной статье.

Анализ исследований

В работах [1-3], посвященных анализу языка и речи, рассматриваются математические модели и методики, предназначенные для ограниченного класса звуков речи (тональных звуков).

Постановка задачи

Провести численное исследование шипящих согласных звуков с целью формирования их эталонов.

Основной материал

По амплитудным и частотным характеристикам все звуки можно разделить на четыре группы (рис. 1).

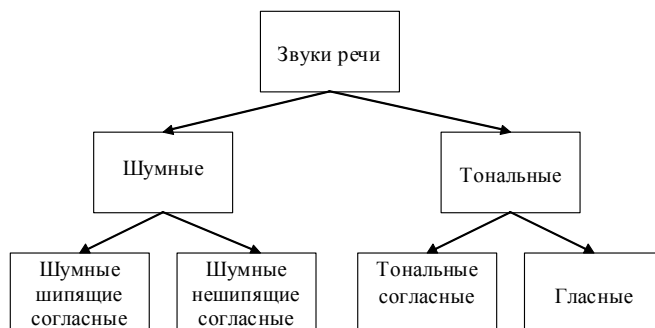


Рисунок 1 – Классификация групп звуков речи

Объектом исследования будут выступать шумные (невокализованные) шипящие звуки. Это объясняется тем, что шипящим звукам уделялось меньше внимания, чем тональным. Этим они представляют интерес для исследования. В качестве признаков для дальнейшей классификации шипящих звуков предлагается использовать положение формантных и антиформантных областей в спектре. Форманта – это частота F_i , которой соответствует максимум амплитудного значения A_i в спектре сигнала. Антиформанта – это частота Φ_i , которой соответствует минимум амплитудного значения B_i в спектре сигнала. Пример формантного представления шипящего звука приведен на рис. 2.

Эталоны шипящих звуков, содержащие форманты и антиформанты, представлены в виде

$$E = ((\bar{F}_1, \bar{A}_1), (\bar{\Phi}_1, \bar{B}_1), \dots, (\bar{F}_n, \bar{A}_n), (\bar{\Phi}_n, \bar{B}_n), (\hat{F}_1, \hat{A}_1), (\hat{\Phi}_1, \hat{B}_1), \dots, (\hat{F}_n, \hat{A}_n), (\hat{\Phi}_n, \hat{B}_n)), \quad (1)$$

где $(\bar{F}_i, \bar{A}_i), (\bar{\Phi}_i, \bar{B}_i)$ – математические ожидания i -х формант и антиформант, $(\hat{F}_i, \hat{A}_i), (\hat{\Phi}_i, \hat{B}_i)$ – среднеквадратичные отклонения i -х формант и антиформант.

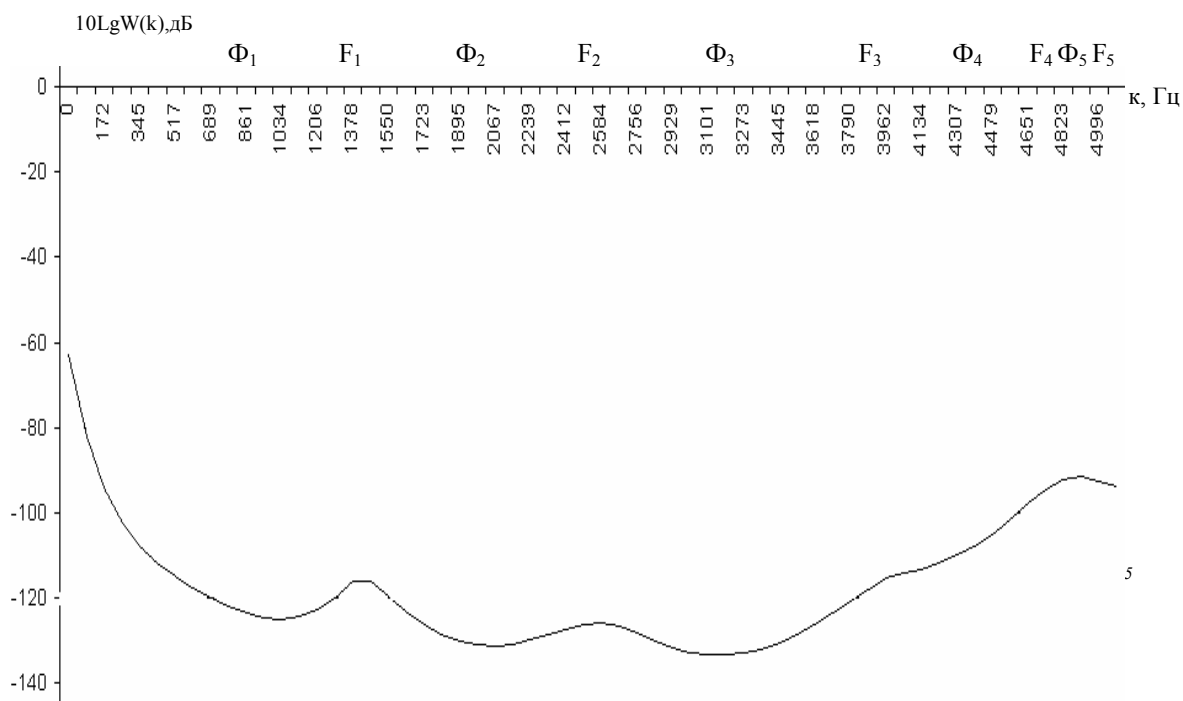


Рисунок 2 – Огибающая спектра для шипящего согласного звука

При выделении частотных признаков согласных звуков использовался подход, основанный на параметрах линейного предсказания [3, р. 507] и состоящий в том, что текущее значение речевого сигнала аппроксимируется линейной комбинацией предшествующих значений сигнала. Согласно этому методу, речь можно представить в виде сигнала на выходе линейной системы с переменными по времени параметрами,

возбуждаемой квазипериодическими импульсами (в пределах тонального сегмента) или случайным шумом (на шумном сегменте).

На рис. 3 изображена модель преобразования речи [3].

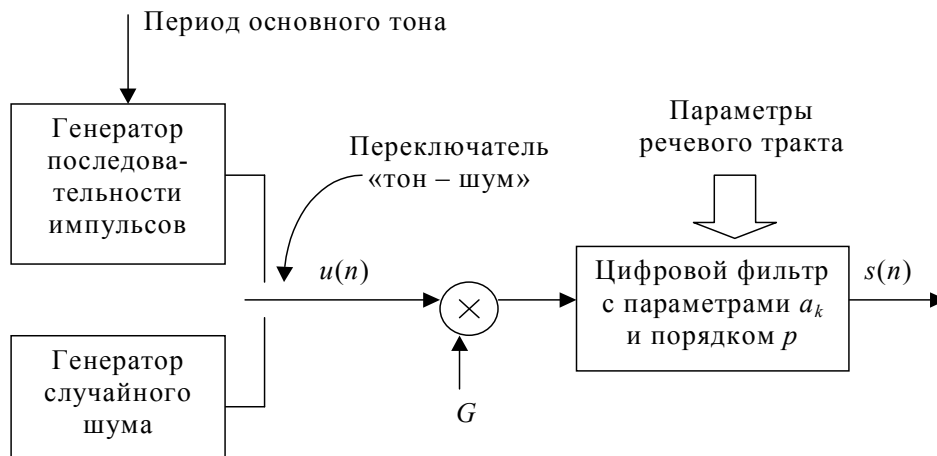


Рисунок 3 – Структурная схема модели преобразования речи

Для модели рис. 3 текущее значение речевого сигнала $s(n)$ связано с сигналом возбуждения $u(n)$ простым разностным уравнением

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n), \quad (2)$$

где G – параметр усиления, a_k – параметры цифрового фильтра, p – порядок цифрового фильтра.

Выходной сигнал цифрового фильтра с параметрами a_k определяется в виде

$$\tilde{s}(n) = \sum_{k=1}^p a_k s(n-k). \quad (3)$$

Погрешность определяется как

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k). \quad (4)$$

Параметры a_k были вычислены согласно алгоритму лестничного метода [3], а параметр усиления G – согласно

$$G^2 = \sum_{n=0}^{N-1} e^2(n), \quad (5)$$

где N – длина анализируемого сегмента

Используя метод линейного предсказания, авторами было проведено численное исследование шипящих согласных звуков, которое заключалось в следующем. Диктор произносил по 100 вариантов каждого звука. Запись производилась с использованием стандартного микрофона с частотными характеристиками 80 – 20000 Гц, после чего производилась оцифровка с частотой дискретизации 22050 Гц и уровнем квантования 8 бит. Полученные

оцифрованные сигналы обрабатывались с использованием выбранных методов с размером окна $N=512$ и половинным перекрытием (т.е. текущий сегмент перекрывает правую половину предшествующего ему сегмента). Таким образом, весь сигнал разбивался на перекрывающиеся сегменты. Для каждого сегмента определялся энергетический спектр $W(k)$ (рис. 4) в частотном диапазоне от 0 до 8000 Гц, при этом $p=30$.

$$W(k) = \frac{G^2}{r(0) - \sum_{n=1}^p r(n) \cos\left(\frac{2\pi}{N} nk\right)}, \quad 0 \leq k \leq \frac{8000 * N}{22050}, \quad (6)$$

где $r(n)$ – автокорреляционная функция,

$$r(0) = \sum_{k=0}^p a_k^2, \quad r(m) = 2 \sum_{k=0}^{p-r} a_k a_{k+r}. \quad (7)$$

Для удаления побочных максимумов энергетический спектр логарифмировался. В сглаженном спектре определялись форманты и антиформанты. Каждая форманта и антиформанта оценивалась по матожиданию и дисперсии (использовалось 100 реализаций).

Результаты численного исследования одного из шипящих звуков (звук [с]) представлены в табл. 1,2.

Таблица 1 – Форманты звука [с]

Стат. Оценки	A ₁	F ₁	A ₂	F ₂	A ₃	F ₃	A ₄	F ₄	A ₅	F ₅
М	-112	1517	-122	2598	-107	3905	-88	4696	-83	4910
σ	14	202	9	203	14	261	9	115	7	0

Таблица 2 – Антиформанты звука [с]

Стат. Оценки	B ₁	Φ ₁	B ₁	Φ ₁	B ₁	Φ ₁	B ₁	Φ ₁	B ₁	Φ ₁
М	-126	1034	-133	2088	-134	3248	-112	4343	-97	4890
σ	11	200	8	190	8	253	12	251	8	38

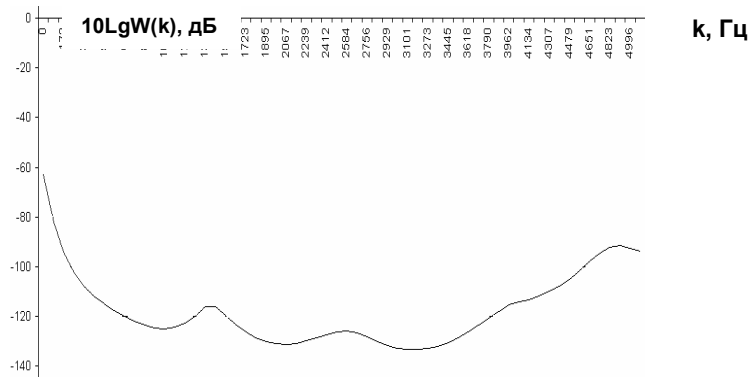


Рисунок 4 – Форма огибающей спектра для шипящего согласного звука [с]

На основании среднестатистических оценок формант и антиформант эталоны (1) заполняются соответствующими значениями для каждого класса шипящих согласных звуков. Например, эталон для звука [с], согласно табл. 1, представлен в виде

$$E = ((1517,-112), (1034,-126), (2598,-122), (2088,-133), (3905,-107), (3248,-134), (4696,-88), (4343,-112), (4910,-83), (4890,-96), (202,14), (200,11), (203,9), (190,8), (261,14), (253,8), (115,9), (251,12), (0,7), (38,8)).$$

Выводы и перспективы

В данной работе с помощью метода линейного предсказания проведено численное исследование шипящих согласных звуков с целью формирования их эталонов. Основные положения данной работы предназначены для реализации в интеллектуальных системах управления, в которых команды отдаются голосом.

Литература

1. Рабинер Л.Р., Шафер П.В. Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981. – 496 с.
2. Шевченко А.И., Федоров Е.Е. Разработка естественно-языкового интерфейса интеллектуальной медицинской системы // Труды Междунар. дистанционной науч.-практ. конф. «Автоматизированное Рабочее Место Врача 2002». – Днепропетровск. – 2002. – С. 327-330.
3. Rabiner L.R., Jang B.H. Fundamentals of speech recognition. – New Jersey: Prentice Hall PTR, Englewood Cliffs, 1993.

Є.Є. Федоров, І.О. Шевцова

Чисельне дослідження шиплячих приголосних звуків

При створенні систем розпізнавання мови важливу роль відіграє вибір ознак, на основі яких формуються еталони звуків мови. Для вирішення цієї задачі у статті наведено чисельне дослідження шиплячих приголосних звуків.

U. Fedorov, I. Sevtsova

Numeral research of voice-frequency vowels

At creation of the systems of speech recognition important part is acted by the choice of signs which the standards of sounds of speech are formed on the basis of. For the decision of this task numeral research of voiced consonantal is resulted in the article.

Статья поступила в редакцию 13.07.2004.