$d_k > a_k^{NZ}; \quad d_{k+1} > a_{k+1}^{NZ}; \quad \beta_k, \beta_{k+1} < 0$

Параметры α_i, α_{i-1} можно определить исходя из следующих соотношений:

$$\frac{\ln\left(1-\frac{b_k^{NZ}}{d_k}\right)}{t_k} \le \beta_k \le \frac{\ln\left(1-\frac{a_k^{NZ}}{d_k}\right)}{t_{k+1}}; \quad \frac{\ln\left(\frac{a_{k+1}^{NZ}}{d_{k+1}}\right)}{t_{k+1}} \le \beta_{k+1} \le \frac{\ln\left(\frac{b_{k+1}^{NZ}}{d_{k+1}}\right)}{t_{k+2}}.$$

При построении результирующей функции будем опираться на тот факт, что их значения не являются ранозначными. Так усталость и отдых в течение дня меньше, чем усталость, накопленная за месяц или год.

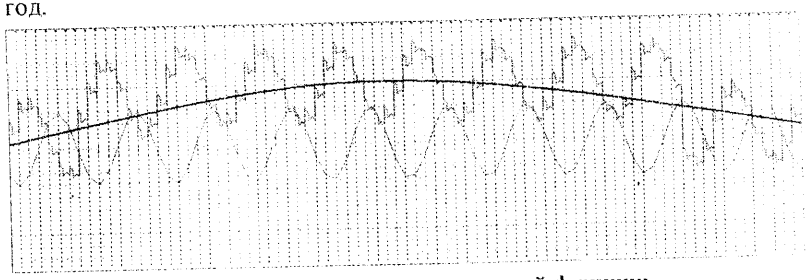

Рис.2. Построение результирующей функции

Результирующую функцию (7) можно построить задавая граничные значения для функций (3)–(6).

Выводы. Значения параметров для функции (7) являются индивидуальными и неизвестны. Принимая во внимание то, что значение (7) не может превышать 1, были сделаны предположения относительно параметров функций (3)–(6).

Оценка надежности распознавания является сложной задачей, требующей учёта множества факторов как функции времени, а не константы. Нужно найти распределение вероятностей надежности в течение длительного периода (не менее нескольких месяцев) и определить все ее взлеты и падения.

1. **Уинфри А.Т.** Время по биологическим часам. / А.Т. Уинфри – М., 1990. – 250с.

Надійшла до редколегії 01.06.09

УДК 534.4: 621.391

М.П. Савельев, О.Н. Карпов

Днепропетровский национальный университет им. Олеся Гончара

СИНТЕЗ РЕЧИ В ФОРМАНТНО-ГОЛОСОВОЙ МОДЕЛИ

Вміщено детальний опис існуючих методів синтезу мови. Описані переваги і недоліки кожного з них. Розглянуті проблемні місця і можливі підходи до їх вирішення. Описано загальний склад синтезатора мови і його компонентів.

Ключові слова: *алофон, каскадна модель, лінійна модель, синтез мови, синтезатор, фонема*

Подробно описаны существующие методы синтеза речи. Описаны преимущества и недостатки каждого из них. Рассмотрены проблемные места и возможные подходы к их решению. Описано общее устройство синтезатора речи и его компонентов.

Ключевые слова: *аллофон, каскадная модель, линейная модель, синтез речи, синтезатор, фонема*

Contains detail description of the modern language synthesis. Advantaged and drawbacks of each are described in details. Problematic places and possible solutions for it are described as well. General form of language synthesis system and all it components are described.

Keywords: *allophone, linear model, waterfall model, language synthesis, synthesis system, phoned.*

Существует много методов реализации формантного синтеза речи. Все они основаны на детальном знании фонем и фонетическом расчленении речи и базируются на двух фундаментальных понятиях: лингвистического – фонемы, и акустического – форманты.

Фонема – основная единица звукового строя языка. Звуковой состав различных языков имеет свои особенности. В русском языке насчитывают 41 фонему, из них 6 гласных и 35 согласных (в английском – 20 гласных и 24 согласных, в французском – 15 гласных и 20 согласных). Можно сказать, что фонема – наименьшая языковая единица, имеющая смыслоразличительное значение. Из последовательности фонем строятся слова. Смысл высказывания выражается посредством цепочки слов.

Под формантами понимаются частотные резонансы (полюса передаточной функции) речевой акустической системы. Параметры формант (частота, ширина, уровень) опеределяются акустическими свойствами системы. Наиболее важный параметр – частота форманты, тесно связан с геометрической конфигурацией речевого тракта. Поскольку в процессе речи конфигурация речевого тракта меняется, то соотвественно меняются формантные частоты (рис. 1).

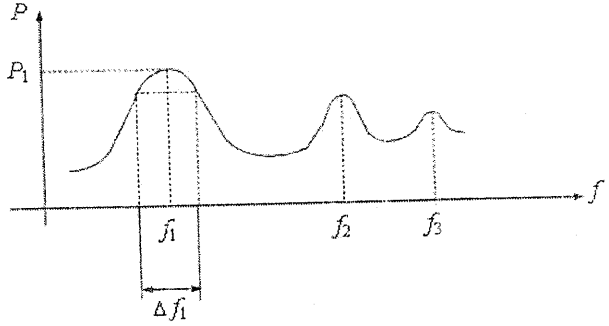

Рис. 1. Пример частотного спектра речи

Для удовлетворительного синтеза речи обычно нужны две – четыре формантные частоты. Они лежат в диапазоне от 200 (первая форманта мужского голоса) до 2000 Гц (третья форманта женского голоса). Точным расположением формантных частот в звуковом спектре и определяется звук, который мы интерпретируем как речь. Причем, все формантные частоты присутствуют в речи одновременно и непрерывно перемещаются вверх-вниз по частотному спектру в соответсвии с особенностями произносимого слова. Поэтому, слушая говорящего человека, вы слышите звук не какой-либо одной частоты, а множество обертонов, которые образуются при фильтрации импульсов, формируемых на выходе голосового тракта.

Итак, в основе формантного синтеза лежит аналогия с моделью речеобразования человека. Рассмотрим формирование гласных звуков на модели (рис. 2).

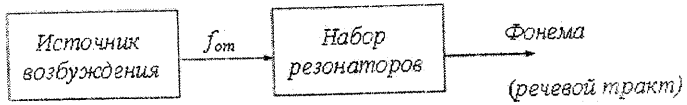

Рис. 2. Модель речеобразования

Источник возбуждения создает импульсы основного тона, частота следования которых непрерывно меняется в процессе формирования

речи. Речевой тракт при образовании гласных звуков работает как набор резонаторов, в которых происходит фильтрация сигнала возбуждения. В результате образуется спектральная картина, содержащая ряд максимумов. Максимумы соответствуют резонансам тракта (это и есть форманты). Таким образом, форманты – это некоторая частотная область концентрации энергии в спектре звука. Используют от двух до шести формант в зависимости от требуемой точности анализа речи. Суммарный выходной сигнал формантных фильтров (резонаторов) достаточно близко соответствует частотному спектру речи человека, и наш слух воспринимает его как речевое сообщение.

Приведем таблицу формантных частот для некоторых фонем гласных звуков.

Таблица 1

Формантные частоты некоторых фонем гласных звуков

Фонема	Формантные частоты		
	F_1	F_2	F_3
о	275	850	2400
и	250	2300	3000
а	575	900	2450

Путем одновременной генерации формантных частот F_1, F_2, F_3 согласно таблицы 1 можно получить гласные звуки.

Структурная схема формантного синтезатора гласных звуков приведена на рисунке 3.

Структурная схема форматного синтезатора гласных звуков включает задающий генератор частоты основного тона, полосовые фильтры, перестраиваемые на формантные частоты, соответствующие синтезируемой фонеме с помощью переменных резисторов $R_1 - R_3$ и сумматор, суммирующий сигналы с трех фильтров. В спектрограмме выходного сигнала этой схемы содержатся три формантные частоты, идентичные формантным частотам в спектрограмме речи человека, произносящего те же гласные.

Гораздо сложнее формировать согласные звуки. Согласные – звуки речи, при произношении которых в полости рта образуются преграды для выдыхаемого воздуха:

- взрывные – при полном смыкании органов речи (п, т, к);
- фрикативные – образуется щель (с, ф, х);
- носовые согласные (н, м);
- аффриката – согласный звук, представляющий слитное сочетание (ч -тш, ц -тс).

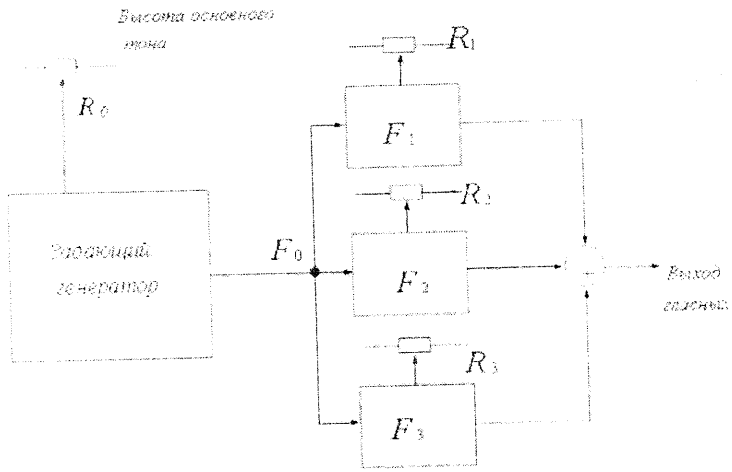

Рис. 3. Структурная схема формантного синтезатора гласных звуков

Чтобы расширить диапазон синтезатора (рис. 3), необходимо ввести источник шума для формирования взрывных и фрикативных согласных, а также аналог носовой резонансной полости, имитирующий носовые согласные. Структура этого расширенного формантного синтезатора приведена на рисунке4.

Структура полного формантного синтезатора речи (рис. 4) усложняется не очень сильно, по сравнению с синтезатором гласных звуков. Значительно увеличилось количество регулировок в схеме. Три из них служат для управления амплитудой фрикативных, гласных и носовых звуков, один – для регулировки высоты тона, а пять остальных – для регулирования частот различных резонансов. Применив в качестве устройства управления регуляторами микропроцессор с соответствующим количеством портов ввода-вывода мы получим устройство, способное производить все необходимые регулировки со скоростью, достаточной для приемлемого приближения к нормальной речи человека.

Естественно, что чем больше обращений к справочной таблице будет производить микропроцессор по каждой фонеме, тем большей плавностью будет отличаться синтетическая речь и тем ближе она будет к естественной человеческой речи.

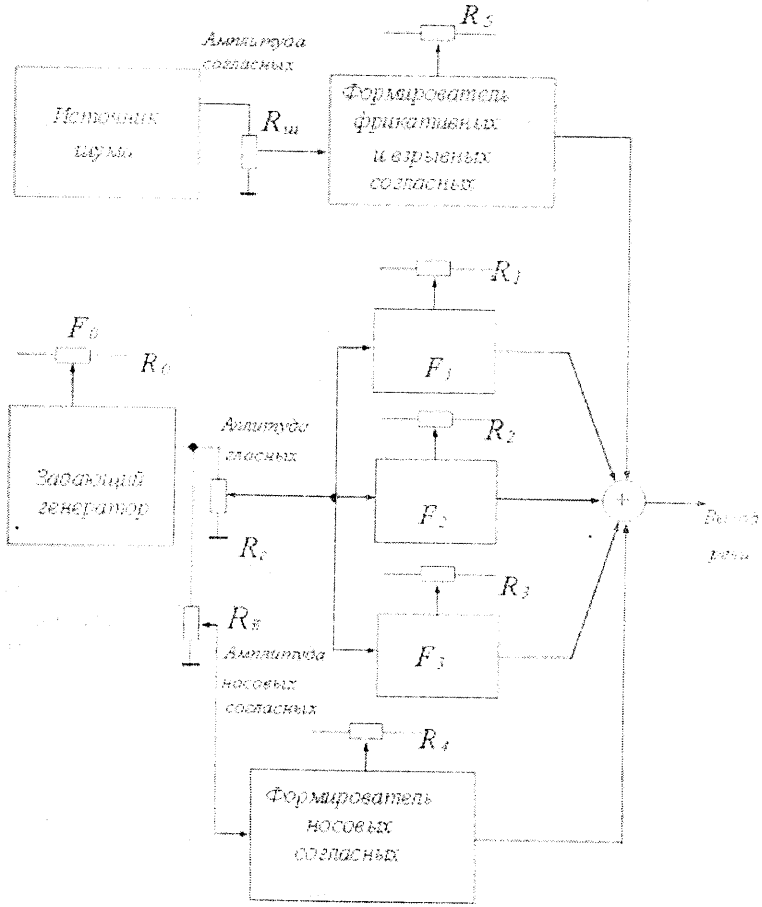

Рис. 4 .Структурная схема формантного синтезатора речи

Преимущество формантного метода синтеза – в его универсальности (т.е. возможность иметь неограниченный словарь) так как здесь речь создается из отдельно генерируемых звуков. Правильно расставив звуки, можно произнести любое слово.

Универсальность эта, однако, не дается бесплатно – за нее приходится расплачиваться ухудшением разборчивости речи. Без соответствующей подготовки трудно понять, что говорит синтезатор.

Дополнительные трудности при реализации большого словаря создает множество имеющихся исключений из правил написания и произношения слов. Если проанализировать одну и ту же фонему,

встречающуюся в различных словах, то может оказаться несколько вариантов произношения данной фонемы. Вариации произносимых фонем называют аллофонами. Аллофоны подразделяются на комбинаторные и позиционные. Комбинаторные оттенки обусловлены соседством данной фонемы с другими фонемами и являются следствием наложения одного звука на другой. Позиционные оттенки обусловлены положением фонемы в слове или фразе по отношению к ударному слогу, концу и началу слова и т.д.

Учет всех факторов позволяет оценить общее число аллофонов, необходимое для качественного синтеза русской речи. Общее число аллофонов гласных $N_A^V = 480$ и согласных $N_A^C = 8800$.

Другой класс лингвистических понятий, учет которых исключительно важен при создании систем синтеза речи, составляют интонация и ударение. Физически интонация и ударение реализуются совокупностью акустических средств (просодикой), к числу которых относятся:

1) мелодика (движение частоты основного тона голоса);
2) ритмика (текущее изменение длительности звуков и пауз);
3) энергетика (текущее изменение силы звука).

Этап преобразования печатного текста в последовательность фонем должен сопровождаться выделением информации, необходимой для задания просодических характеристик синтезируемых речевых сигналов.

Для этой цели текст анализируется и по определенным правилам разбивается на основные единицы: фраза, синтагма, акцентная группа, фонетическое слово.

Эти единицы маркируются, соответственно фразовым, синтагматическим, групповым и словесным ударениями. Каждой синтагме присваивается один из возможных интонационных типов. Это завершенность, незавершенность, вопрос или восклицание.

Под синтагмой понимают слово (или группу слов), представляющее собой цельную синтаксическую интонационно-смысловую единицу. Пример синтагмы представлен на рис. 5.

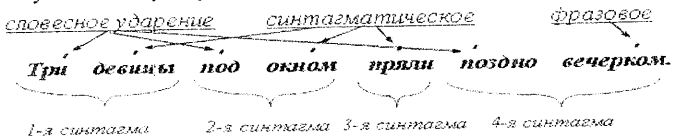

Рис. 5. Пример синтагмы

Таким образом, в качестве входной информации текстового сообщения используется размеченный орфографический текст, то есть

обычный орфографический текст с проставленными знаками словесного, синтагматического и фразового ударений

Эта модель может быть реализована с применением нейронных сетей и допускает самообучение. К сожалению, ввиду сложности точного моделирования особенностей речевого тракта, а также учета интонационной модуляции речи формантно-голосовая модель обладает относительно низкой точностью синтезируемых звуков речи. Тем не менее, современные программы синтеза речи, построенные с использованием этой модели, синтезируют вполне разборчивую речь и могут применяться в ряде случаев.

Заметим, что системы голосового предупреждения о возникновении аварийных ситуаций лучше строить с использованием модели компилятивного синтеза, так как разборчивость речи в таких системах выходит на передний план.

Что же касается «бытовых» синтезаторов речи, то в них можно с успехом применять и форманто-голосовую модель. Схематически эта модель показана на рисунке 6 [3].

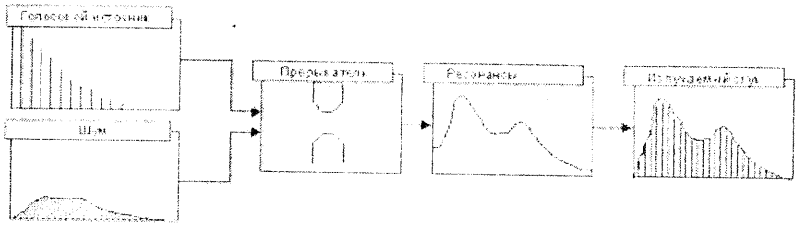

Рис. 6. Формантно-голосовая модель синтеза речи

При построении модели в [3] использовались данные об артикуляционном аппарате человека, а также данные фонетики и лингвистики. Как видите, в качестве исходного сигнала применяется комбинация голосового источника и генератора шума. Прерыватель и резонансное устройство моделирует работу речевого тракта. В результате этого моделирования образуется излучаемый звук речи.

При этом для достижения компромисса между качеством модели и ее сложностью были выбраны следующие основные параметры исследуемой системы:

- частота основного тона;
- частота шума;
- количество формант;
- центральная частота каждой форманты;
- вклад каждой форманты.

Частота основного тона определяет высоту голоса. Этот параметр не должен вызывать у Вас никаких вопросов. Что же касается частоты шума, то здесь нужно сделать пояснение.

Как замечает автор работы [3], образование шума представляет собой достаточно сложный процесс, зависящий от многих факторов, таких как давление и скорость воздушной струи, геометрической формы воздушного тракта, акустических свойств материала и пр. Чтобы полностью смоделировать шум речи на физическом уровне, необходимо создать точную модель речевого аппарата человека, что представляет собой очень сложную задачу.

В качестве альтернативы автор работы [3] использует белый шум, спектр которого распределен по некоторому закону (например, по Гауссу) относительно некоторой центральной частоты. При этом закон распределения подбирается экспериментально, а частотой шума в этом случае является упомянутая выше центральная частота.

Количество активных формант, участвующих в образовании речи, выбирается в [3] экспериментально, причем в качестве ориентировочного значения используется 4.

Так как форманта представляет собой резонанс в речевом тракте, у неё есть частота резонанса и огибающая. Вид огибающей также определяется экспериментально, в первом приближении это Гауссово распределение.

Вклад каждой форманты определяет, насколько сильно форманта воздействует на основной сигнал.

Все приведенные выше параметры, кроме количества формант, изменяются в процессе образования речи для получения различных звуков. Хотя для более качественного синтеза речи необходимо строить более детальную модель, приведенные в [3] параметры достаточны для того, чтобы синтезированные звуки были разборчивы.

Библиографические ссылки

1. **Бондарко Л.В.** Звуковой строй современного русского языка. / Л.В. Бондарко. – М., 1997.
2. **Мерцалова Г.Н.** Лекции по языкознанию (http://www.tula.net/tgpu/resources/yazykozn/index.htm). / Г.Н. Мерцалова // Тульский государственный педагогический университет им Л.Н. Толстого.
3. **Москаленко А.М.** Использование нейросетей для анализа звуковой информации (http://alexmoshp.chat.ru/index.htm). / А.М. Москаленко // Дипломная работа. Кубанский государственный университет.

4. **Алексеев В.** Услышь меня, машина. / В. Алексеев // Компьютерра, №49, 1997.
5. **Захаров Л.** Проблемы создания аллофонной базы автоматического синтеза речи (http://art.bdk.com.ru/govor/rasp.htm).
6. **Панов М.В.** Русский язык. История русского литературного языка. / М.В. Панов // Еженедельник «Русский язык», №26, 2002.
7. **Ундриц В.Ф.** Болезни уха, горла и носа (руководство для врачей). / В.Ф. Ундриц, К.Л. Хилов, Н.Н. Лозанов, В.К. Супрунов. // Медицина, 1969.
8. **Бекеши Г.** Механические свойства уха. / Г. Бекеши, В.А. Розенблит. – М., 1963.
9. **Хоровиц П.** Искусство схемотехники: В 2-х т. Пер. с англ. / П. Хоровиц, У. Хилл. – М, 1984.
10. **Фролов А.В.** Мультимедиа для Windows. Библиотека системного программиста. т. 15 (http://info.datarecovery.ru) / А.В. Фролов, Г.В. Фролов. – М., 1994.
11. **Уоссермен Ф.** Нейрокомпьютерная техника: Теория и практика. / Ф. Уоссерман. – М, 1992.
12. **Головко В.А.** Нейронные сети: обучение, организация и применение. / В.А. Головко. – М., 2001.
13. **Галушкин А.И.** Нейрокомпьютеры. / А.И. Галушкин. – М., 2000.
14. **Круглов В.В.** Искусственные нейронные сети. Теория и практика. / В.В. Круглов, В.В. Борисов. – М., 2002.
15. **Медведев, В.С.** Нейронные сети. MATLAB 6. / В.С. Медведев, В.Г. Потемкин. – М., 2002.
16. Speech Analysis FAQ. (http://svr-www.eng.cam.ac.uk/~ajr/SA95/SpeechAnalysis.html).
17. **Куссуль Э.М.** Ассоциативные нейроподобные структуры. / Э.М. Куссуль. – Киев, 1990.
18. **Нуссбаумер Г.** Быстрое преобразование Фурье и алгоритмы вычисления сверток. / Г. Нуссбаумер – М., 1985.
19. **Астафьева Н.М.** Вейвлет-анализ: основы теории и примеры приведения. Успехи физических наук. / Н.М. Астафьева. – М., 1996. – Т. 166. – № 11.

Надійшла до редколегії 09.07.09