Об одном алгоритме оценки формантных частот на интервале сомкнутых голосовых складок

Конев А.А.
Мещеряков Р.В.
Жевуров С.В.
Хлебников В.С.

Омский государственный университет систем управления и радиоэлектроники
634050 г. Томск, пр. Ленина, 40.
Тел. (факс) (3822) 413-426. E-mail: office@keva.tusur.ru

В статье рассматривается подход к сегментации вокализованных участков речевого сигнала на два класса звуков. В первый класс звуков входят звонкие согласные, во второй класс — сонорные. Подход к сегментации основан на анализе сигнала после одновременной маскировки, т.е. на особенностях восприятия различных классов звуков слуховой системой человека. Представлена структура сегментов сигнала после одновременной маскировки, соответствующих различным классам звуков, и обозначены основные отличия. Рассмотренные отличия могут быть использованы при создании алгоритма автоматической сегментации вокализованных участков сигнала на предложенные классы.

В теории речеобразования рассматриваются два типа источников речевого сигнала — голосовой и шумовой [1]. Голосовой источник генерирует квазипериодический сигнал, характеризующийся наличием гармонической структуры. В сигнале, генерируемом шумовым источником, гармоническая структура отсутствует либо является слабовыраженной.

Кроме сигналов, генерируемых исключительно голосовым или только шумовым источником, речеобразующая система человека способна генерировать сигналы, в образовании которых могут участвовать одновременно оба типа источника.

С другой стороны, в фонетике существует классификация звуков речи, учитывающая тип источника, сгенерировавшего звук [2]. В соответствии с этой классификацией к звукам, образованным с использованием только голосового источника, относятся сонорные, с использованием только шумового — глухие согласные, с использованием обоих источников — звонкие согласные.

Таким образом, подобная классификация существует как со стороны речеобразования, так и со стороны речевосприятия. Значит, слуховая система должна воспринимать параметры речевого сигнала, позволяющие различить соответствующие классы звуков.

Слуховая система человека обладает эффектом одновременной маскировки частот [3]. Он возникает в том случае, когда рядом расположенные нейроны воспринимают две или более компоненты, частоты которых находятся недалеко друг от друга. При этом частота с более высокой амплитудой подавляет частоту с более низкой амплитудой, вплоть до того, что вторая частота может вообще не восприниматься.

Используемая в исследованиях система фильтров основана на модели периферической части слуховой системы человека [4]. В данной модели учитывается эффект одновременной маскировки. После одновременной маскировки сигнал имеет структуру, представленную на рис.1.

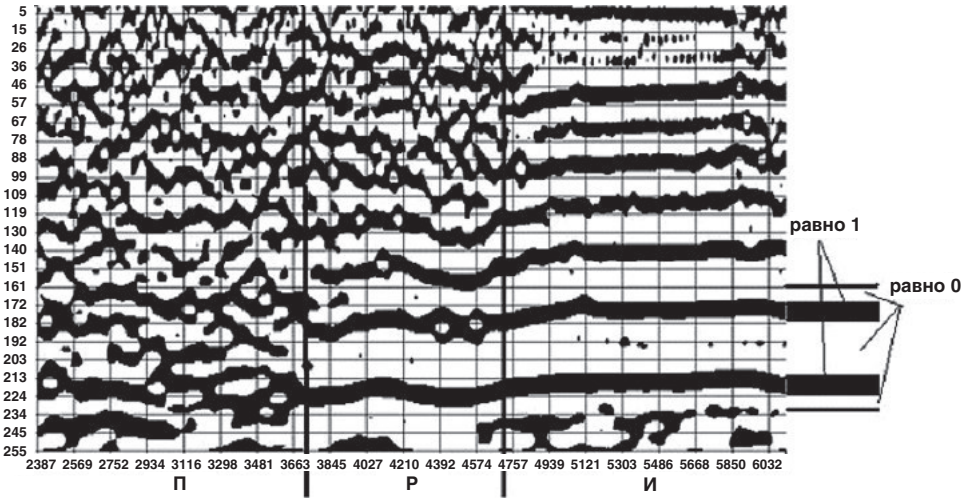

Рис. 1. Структура речевого сигнала после одновременной маскировки

Изображённая структура представляет собой набор бинарных данных. По оси абсцисс — время в дискретных отсчётах (частота дискретизации — 8 кГц), по оси ординат — номера частотных каналов фильтрации (0 канал — 2500 Гц, 255–50 Гц). Чёрным цветом выделены компоненты, которые представлены значением на частотном канале равном единице. Эти компоненты воспринимаются слуховой системой человека. Белым цветом выделены компоненты, невоспринимаемые слуховой системой и которые представлены значением на частотном канале, равном нулю. На сегменте, соответствующем звуку «и», чётко просматривается «полосатая» гармоническая структура сигнала. В «полосу» входят сама гармоника и прилегающие к ней незамаскированные частотные компоненты. Таким образом, «полоса» — непрерывный интервал единиц на одном временном отсчете. На сегменте, соответствующем звуку «п», гармоническая структура отсутствует.

На основе полученной структуры был создан алгоритм сегментации сигнала на вокализованные и невокализованные участки, описанный в [5]. Разработанный алгоритм в автоматическом режиме сегментирует речевой сигнал на вокализованные и невокализованные сегменты с надёжностью более 90%. При этом вокализованные участки включают в себя звонкие согласные и сонорные звуки, а невокализованные — глухие согласные и «тишину».

Алгоритм основан на анализе в каждый дискретный момент времени частотной области, включающей две гармоники речевого сигнала (два непрерывных интервала единиц определённой длины, разделённых интервалом нолей). Для работы алгоритма создаётся набор шаблонов, с которыми сравнивается структура сигнала в текущий момент времени. Каждый шаблон состоит из эталонной последовательности нолей и единиц, характеризующих вокализованный сигнал с определённой частотой основного тона. Пример шаблона приведён в правой части рис. 1.

В данном исследовании был проведён анализ возможности применения аналогичного подхода для сегментации вокализованных сегментов на участки, соответствующие сонорным и звонким согласным звукам. Исследование заключалось в визуальном анализе структуры сигнала после одновременной маскировки на участках, соответствующих различным классам вокализованных звуков. При визуальном анализе оценивалось наличие/отсутствие гармонической структуры сигнала на каждом участке.

Структура речевого сигнала после одновременной маскировки приведена для женского (рис. 2) и мужского (рис. 3) голосов. В качестве примера представлен участок слова «предложил».

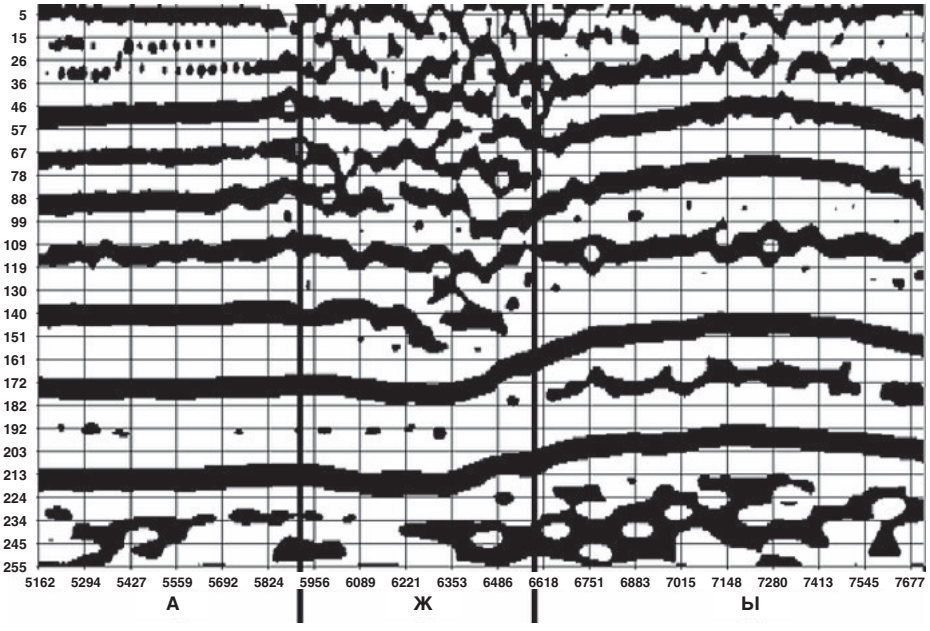

Рис. 2. Структура сонорных и звонких согласных звуков (диктор-женщина)

После проведения визуального анализа было установлено, что гармоническая структура у звонких согласных чаще всего отсутствует на частотах выше 2–4 гармоники. По предварительным исследованиям гармоническая структура у более 80% звонких согласных отсутствует на частотах выше 800–900 Гц. Примерно в половине случаев гармоническая структура отсутствует на частотах выше 500 Гц.

Гармоническая структура у сонорных прослеживается до 7–8 гармоники. Часто промежуточные «полосы» 3–5-й гармоник отсутствуют частично или полностью (например, 6-я гармоника у звука [а] на рис. 2). При этом гармоники, расположенные на частотах выше частоты отсутствующей гармоники, хорошо прослеживаются. По предварительным исследованиям у 90% сонорных существует гармоническая структура на частотах выше 800 Гц. Основная часть ошибок у сонорных приходится на безударные гласные и сонанты [р] и [й].

Различие в структуре участков сигнала, соответствующих звонким согласным, сонорным и глухим звукам, является подтверждением адекватности используемой модели периферической части слуховой системы человека. Проведённые исследования показывают связь между артикуляционными

принципами формирования речевого сигнала и особенностями его восприятия. Одновременная маскировка позволяет слуховой системе получить параметры, необходимые для первичной сегментации и классификации речевого сигнала. Что, в свою очередь, даёт возможность использования различных алгоритмов анализа сигнала для различных классов звуков. Например, анализ частоты и интенсивности гармоник можно использовать для анализа звонких согласных (на нижних частотах) и сонорных звуков.

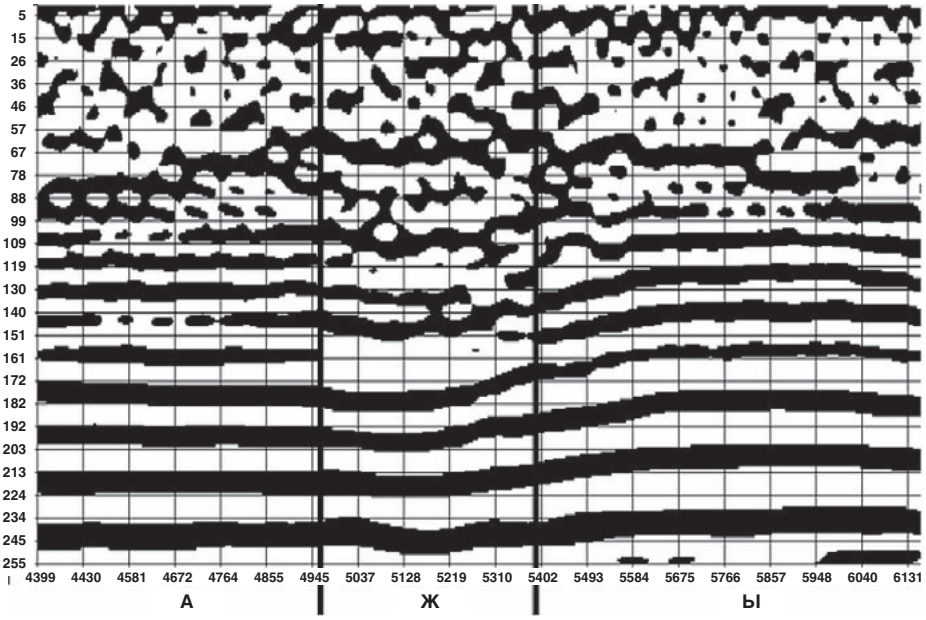

Рис. 3. Структура сонорных и звонких согласных звуков (диктор-мужчина)

Дальнейшие исследования будут направлены на формирование статистики по структуре речевого сигнала на участках звуков различных классов. Данная статистика необходима для выработки требований к алгоритму сегментации вокализованных участков сигнала на сонорные и звонкие согласные.

ЛИТЕРАТУРА

1. *Сапожков М.А*. Речевой сигнал в кибернетике и связи. М.: Государственное издательство литературы по вопросам связи и радио, 1963. 450 с.

2. *Буланин Л. Л*. Фонетика современного русского языка. М.: Высшая школа, 1970. 206 с.

3. Слуховая система / Под ред. Я. А. Альтмана. Л.: Наука, 1990. 620 с.

4. *Bondarenko V. P., Moor V. R., Chabanets A. N*. The analysis of speech perception mechanisms on the models of auditory system // Proceedings XIth ICPhS. Tallinn, 1987. V. 2. P. 77–80.

5. *Конев А. А. Мещеряков Р. В*. Алгоритм сегментации речевого сигнала на вокализованные и невокализованные участки // Сборник трудов XIX сессии Российского акустического общества. Т. III. М.: ГЕОС, 2007. С. 56–60.