

Параметрическое описание ударных гласных звуков

Конев А.А.
Мещеряков Р.В.
Тиунов С.Д.
Черных Д.В.
Чижевская С.Ю.

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

В настоящее время основными параметрами гласных звуков принимаются частота основного тона и форманты. В настоящей работе предлагается определение значения частоты и интенсивности двух максимальных по интенсивности гармоник, выделяемых в областях до 800 Гц и от 800 до 2400 Гц. Кроме того, показывается структура фонем/аллофонов для звукового строя языка. Сформированы критерии структурированного описания гласных звуков. Таким образом, показывается возможность параметрического описания гласных звуков с учётом звукового строя языка.

Наиболее распространённым подходом к построению параметрического описания речевого сигнала является применение одного и того же алгоритма для всех классов звуков. Это упрощает предварительную обработку сигнала, исключает использование предварительной сегментации. С другой стороны, подобный подход не учитывает особенностей системы речеобразования и генерации различных классов звуков.

Предлагаемый подход основан на применении различных алгоритмов получения параметров речевого сигнала для разных классов звуков. Это связано с тем, что вокализованные участки обладают квазипериодической структурой из-за участия голосового источника в генерации речевого сигнала [1]. Квазипериодическая структура позволяет в качестве основы параметрического описания вокализованных звуков использовать параметры гармонических составляющих сигнала. Таким образом, алгоритмы получения параметров сегментов сигнала будут отличаться в зависимости от типа источника (голосовой или шумовой).

Звуки могут быть образованы как при помощи только голосового (гласные, сонанты) или шумового (глухие согласные) источника, так и с использованием обоих типов источников (звонкие согласные). Поэтому для проверки алгоритмов, предназначенных для квазипериодических сигналов, целесообразно проводить анализ звуков, генерируемых только голосовым источником. Из выбранных классов звуков наиболее изученными являются ударные гласные звуки. Чаще всего алгоритмы получения их параметрического описания основаны на анализе формантной структуры [2].

Основываясь на формантной структуре ударных гласных, можно предположить, что в области формантных частот должны находиться максимальные по интенсивности гармоники сигнала. Тогда в параметрическое описание должны быть включены максимальные по

интенсивности гармоники. При этом их поиск должен проводиться в областях первой и второй формант. Граница разбиения всей анализируемой частотной области на области первой и второй формант расположена около 850 Гц [3]. В данном исследовании граница была выбрана равной 800 Гц. В каждой из полученных частотных областей были выбраны по две гармоники с максимальной интенсивностью.

Для получения параметров речевого сигнала, необходимых для описания ударных гласных звуков, использовался программный комплекс, реализующий модель периферической части слуховой системы человека. Этапы обработки сигнала с использованием программного комплекса:

- предварительная фильтрация, позволяющая получить параметры речевого сигнала с высокой точностью (погрешность вычисления частоты основного тона менее 1%) [4];
- одновременная маскировка, позволяющая выделить значимые для слуховой системы компоненты речевого сигнала [5];
- сегментация речевого сигнала на вокализованные и невокализованные участки (надёжность сегментации — более 90%) [6];
- определение каналов фильтрации, соответствующих гармоникам вокализованных участков речевого сигнала и определение значений частоты и интенсивности этих гармоник на каждом дискретном отсчете времени.

Для обработки использовались вручную отсегментированные речевые сигналы дикторов обоего пола. Сигналы содержали слитную речь, например, фразу: «Быть может, от этого именно удара погиб материк, лежавший на запад от Африки в Атлантическом океане».

Для параметрического описания ударных гласных использовались следующие параметры:

- признак вокализованности сегмента;
- частота максимальной по интенсивности гармоники, лежащей в области частот до 800 Гц;
- частота второй по интенсивности гармоники, лежащей в области частот до 800 Гц;
- частота максимальной по интенсивности гармоники, лежащей в области частот от 800 до 2300 Гц;
- частота второй по интенсивности гармоники, лежащей в области частот от 800 до 2400 Гц.

На рис. 1–6 изображена динамика изменения частоты (ось ординат) гармоник, выбранных для параметрического описания ударных гласных, во времени (ось абсцисс, частота дискретизации 12 кГц). При этом чёрным цветом указаны максимальные по интенсивности гармоники в каждой из рассматриваемых областей, а серым вторые по интенсивности гармоники. В нижней части рисунков изображена динамика изменения частоты основного тона на анализируемом участке.

Из рис. 1–2 следует, что на стационарном участке гласной [А] максимальная гармоника в первой области равна 550–700 Гц, т.е. соответствует известным данным по частоте первой форманты. Несмотря на более высокое значение частоты максимальной гармоники во второй области по сравнению со среднестатистической частотой второй форманты, звук [А] можно однозначно определить по значению, соответствующему первой форманте.

Параметры звука [О] диктора-мужчины соответствуют известным значениям частоты первой (400–500 Гц) и второй (800–1100 Гц) формант. Только на переходном к звуку [ж] участке происходит превышение значения частоты второй форманты. У диктора-женщины первая форманта находится между двумя гармониками, что приводит к изменению на протяжении звука [О] номера максимальной гармоники.

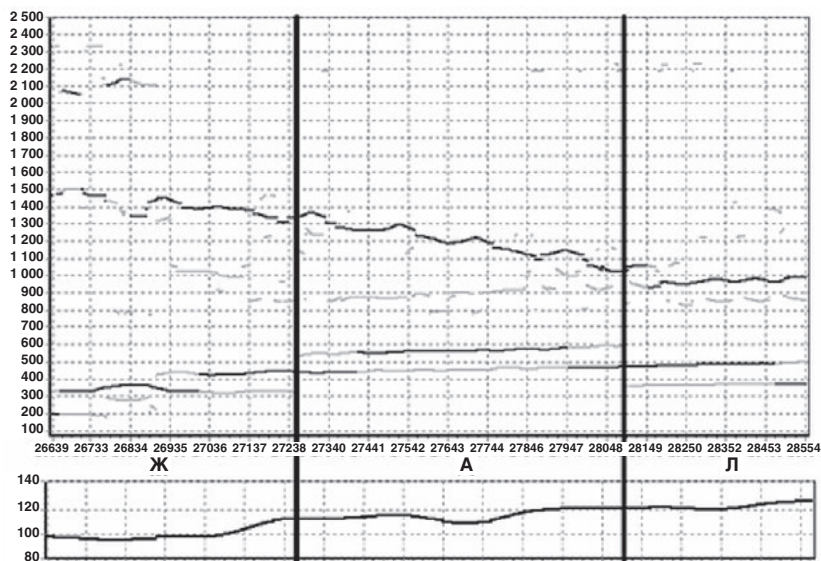


Рис. 1. Звук [А] из слова «жалованье» (диктор — мужчина)

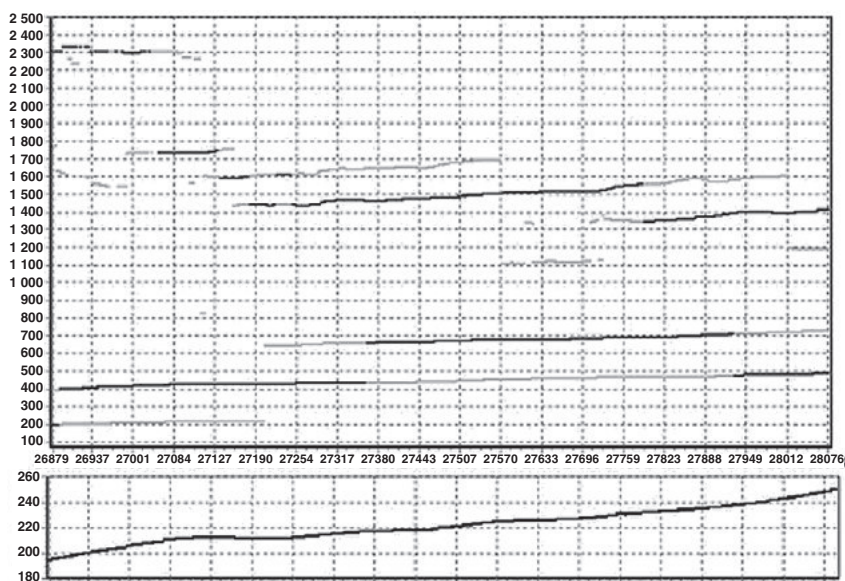


Рис. 2. Звук [А] из слова «жалованье» (диктор — женщина)

Параметры звука [И] обоих дикторов в целом соответствуют известным значениям частоты первой (200–350 Гц) и второй (1800–2200 Гц) формант. Превышение значения частоты второй форманты (до 2400 Гц) связано с мягкостью следующего звука [м].

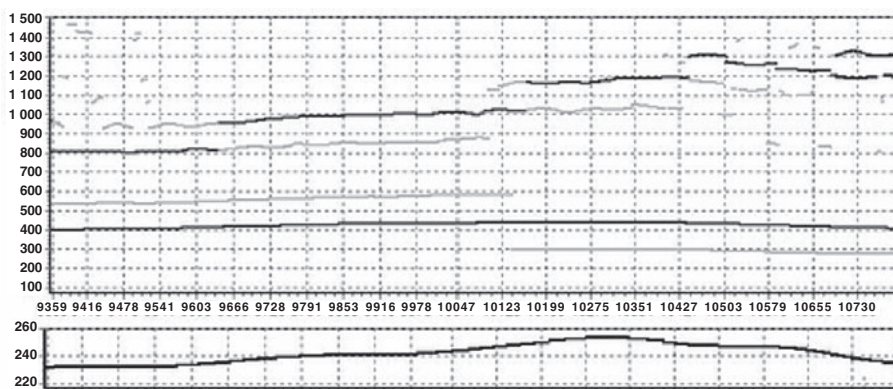


Рис. 3. Звук [O] из слова «может» (диктор — мужчина)

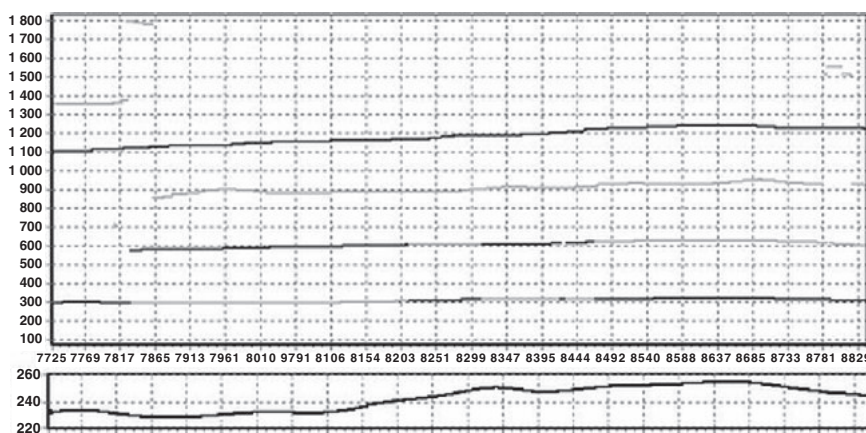


Рис. 4. Звук [O] из слова «может» (диктор — женщина)

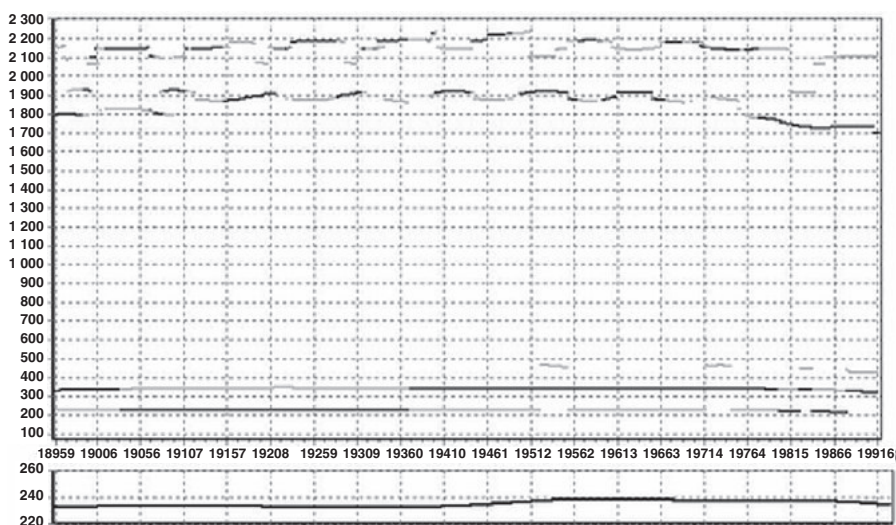


Рис. 5. Звук [I] из слова «именно» (диктор — мужчина)

При анализе данных, полученных по ударным гласным звукам, был выявлен ряд недостатков. В некоторых позициях в слоге сближаются параметры [О] и [У], [А] и [О], [Э] и [Ы], т.е. необходим анализ параметров гласного с учётом окружающих звуков. При пересечении гармоник частотной границы между областями (800 Гц) возможно резкое изменение номера второй по интенсивности гармоники в области. Более высокий уровень интенсивности у низкочастотных гармоник может приводить к некорректному определению максимальных гармоник на некоторых участках (например, на рис. 6 во второй области на некоторое время максимальной становится гармоника с частотой 1150 Гц).

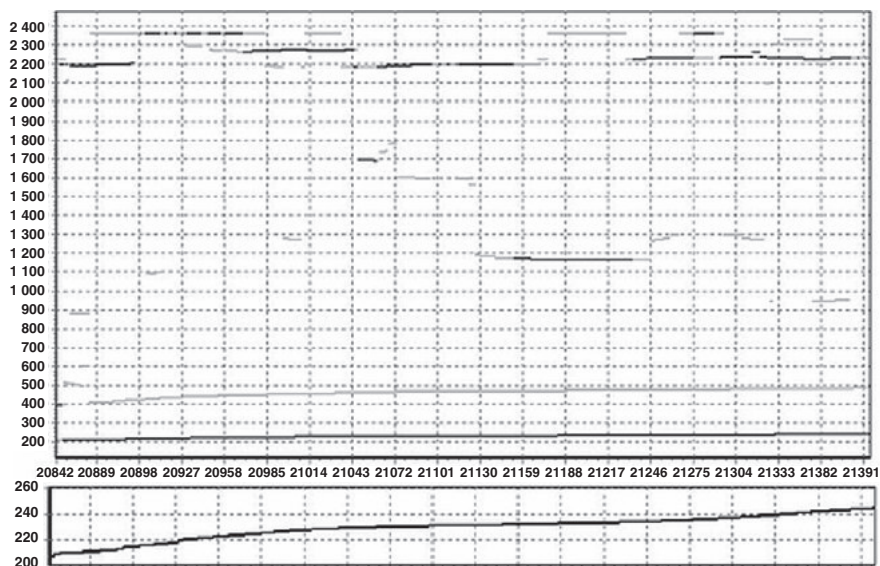


Рис. 6. Звук [И] из слова «именно» (диктор — женщина).

Адекватность предложенного подхода подтверждается совпадением используемых параметров с известными частотами формант гласных звуков. Основное направление дальнейшей работы — набор статистики по ударным гласным звукам, включающей предлагаемые параметры. На основе предложенного подхода планируется создание алгоритма, позволяющего классифицировать ударные гласные, составление параметрического описания безударных гласных и сонант. Кроме того, планируется изучение перехода максимума интенсивности с одной гармоники на другую в качестве дополнительного параметра сегментации вокализованных участков речевого сигнала.

ЛИТЕРАТУРА

1. Сапожков М.А. Речевой сигнал в кибернетике и связи. М.: Государственное издательство литературы по вопросам связи и радио, 1963. 450 с.
2. Huang Xuedong. Spoken language processing: a guide to theory, algorithm and system development. New Jersey: Prentice Hall PTR, 2001. 980 с.
3. Михайлов В. Г., Златоустова Л. В. Измерение параметров речи / Под ред. М.А. Сапожкова. М.: Радио и связь, 1987. 168 с.
4. Бондаренко В. П., Конев А. А. Оценка точности определения значения частоты основного тона речевого сигнала // Сборник трудов XIX сессии Российского акустического общества. Т. III М.: ГЕОС, 2007. С. 33–36.
5. Бондаренко В.П., Пономарев А.А., Рогозинская Е.А. Модель одновременной маскировки // Интеллектуальные системы в управлении, конструировании и образовании Томск: STT, 2004. 216 С. 167–174.
6. 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.