УДК 621.396: 534.78

ФАЗОВЫЕ СООТНОШЕНИЯ МЕЖДУ ОСНОВНЫМ ТОНОМ И ОБЕРТОНАМИ ГЛАСНЫХ ЗВУКОВ

В.И. ВОРОБЬЕВ, Г.В. ДАВЫДОВ, Ю.В. ШАМГИН

Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь,

Поступила в редакцию 20 мая 2005

Путем обработки цифровых записей гласных звуков речи нескольких дикторов показано, что оценки разностей фаз между основным тоном речевого сигнала и обертонами несут информацию, которая может быть использована при решении задач распознавания звуков речи и идентификации дикторов.

Ключевые слова: разность фаз кратночастотных квазигармонических компонентов речевых сигналов, распознавание речи, идентификация дикторов.

Введение

Известно [1], что сдвиги фаз между гармоническими составляющими речевых сигналов (РС) непосредственно не оказывают заметного влияния на их слуховое восприятие. Вместе с тем это не означает, что эти параметры всякий раз приобретают произвольные значения и являются заведомо неинформативными. Напротив, в речевых сигналах, формируемых единой материальной системой и характеризующихся определенной цельностью, неизбежно присутствует внутренняя (в том числе и фазовая) согласованность частотных составляющих.

Наши исследования [2–6] показали, что одним из перспективных видов анализа межкомпонентных связей у сверхширокополосных и полигармонических радио- и гидролокационных сигналов, а также акустических сигналов речи и вибрации в механических узлах машин является анализ разностей фаз кратночастотных составляющих и составляющих с рациональными отношениями частот.

До недавнего времени выявление и анализ связей между компонентами спектра РС были весьма затруднены. Современные средства обработки РС существенно изменяют положение.

На возможную целесообразность учета начальных фаз узкополосных составляющих вокализованных участков РС и формантных колебаний указывалось в [7, 8]. Однако данные по вопросам межкомпонентной фазовой обработки РС в известных нам публикациях не содержатся.

Приведенные ниже материалы наглядно свидетельствуют, что оценки разностей фаз между колебанием с частотой основного тона (ЧОТ) и обертонами гласных звуков несут информацию, которая может быть использована при решении задач распознавания звуков речи и идентификации дикторов.

Математический анализ

Реализации обрабатываемого гласного звука $x(t)$ можно представить в виде

$$x_k(t) = \sum_{p=1}^{p=N} A_{kp}(t)\cos\left(2\pi F_{k0} pt + \Phi_{kp}(t)\right) \quad k = \overline{1, M}\,, \tag{1}$$

где $A_{kp}(t)$ и $\Phi_{kp}(t)$ — медленно меняющиеся амплитуда и фаза *p*-й квазигармонической составляющей для *k*-й реализации звука *x*(*t*); F_{k0} — ЧОТ в *k*-й реализации; *M* — количество реализаций; *N* — число выбранных для анализа квазигармонических составляющих.

В формуле (1) аргумент косинуса представляет собой текущее значение полной фазы *p*-го квазигармонического колебания в *k*-й реализации звука *x*(*t*), равное

$$\Psi_{kp}(t) = 2\pi F_{k0} pt + \Phi_{kp}(t);\ p = \overline{1, N};\ k = \overline{1, M}\,. \tag{2}$$

Если $\Psi_{kp}(t)$ разделить на *p* и результат вычесть из полной фазы $\Psi_{k1}(t)$ колебания с ЧОТ (*p*=1), то определенная таким образом разность фаз $\Delta\Psi_{k1}^{p}(t)$ между колебанием с ЧОТ и *p*-й квазигармонической составляющей не содержит линейно нарастающих слагаемых:

$$\Delta\Psi_{k1}^{p}(t) = \Phi_{k1}(t) - \Phi_{kp}(t)/p\,;\ p = \overline{1, N};\ k = \overline{1, M}\,. \tag{3}$$

Необходимо отметить, что для взаимного уничтожения упомянутых составляющих в формуле (3) требуется обеспечить непрерывность функций $\Psi_{k1}(t)$ и $\Psi_{kp}(t)$, что достигается применением известной процедуры их "сшивания" в точках квазипериодически возникающих в них скачков на величину 2π.

Можно убедиться, что диапазоном однозначного определения величины $\Delta\Psi_{k1}^{p}(t)$ является отрезок [0**;** 2π/*p*]. Поэтому вычисляемые по формуле (3) значения $\Delta\Psi_{k1}^{p}(t)$ необходимо нормировать по модулю $|2\pi/p|$**:**

$$\Delta\Psi_{k1}^{p}(t) = \left[\Phi_{k1}(t) - \Phi_{kp}(t)/p\right] \quad \left|\frac{2\pi}{p}\right. \qquad p = \overline{1, N};\ k = \overline{1, M}\,. \tag{4}$$

В случаях, когда анализируется разность фаз между *q*-й и *p*-й квазигармонической составляющими, формула (3) перепишется в виде

$$\Delta\Psi_{kp}^{p}(t) = \left[\Phi_{k1}(t) - \Phi_{kp}(t)/p\right] \quad \left|\frac{2\pi p}{q}\right. \qquad p = \overline{2, N};\ q = \overline{3, N};\ k = \overline{1, M}\,. \tag{4}$$

Полные фазы $\psi_{kp}(t)$ отфильтрованных и доступных для преобразований по отдельности квазигармонических компонентов $A_{kp}(t)\cos\left(\Psi_{kp}(t)\right)$, $p = \overline{1, N}$, $k = \overline{1, M}$ можно определять с помощью перехода к соответствующим им аналитическим сигналам с использованием преобразования Гильберта.

Методика

Для выполнения оценок разности фаз между квазигармонической компонентой с ЧОТ и обертонами использовалась следующая последовательность действий:

ввод цифровых записей звуков речи;

обработка анализируемой реализации $x_k(t)$, $k = \overline{1, M}$, звука $x(t)$ временным окном Хэннинга с длительностью, равной длительности этой реализации;

вычисление спектра и кепстра мощности реализации $x_k(t)$, $k = \overline{1, M}$;

оценка усредненного на длительности анализируемого звука значения частоты F_{k0} его основного тона, выполняемая на основе данных спектрального и кепстрального анализа;

вычисление средних частот ближайших обертонов $p\,F_{k0}$; p=2, 3, …, N;

полосовая фильтрация звука на ЧОТ (F_{k0}) и на частоте р-го обертона ($p\,F_{k0}$) в полосе (0,1–0,2) от центральных частот;

формирование *аналитических* представлений квазигармонических колебаний на ЧОТ и частоте *p*-го обертона;

вычисление огибающих и фаз колебаний на ЧОТ и частоте *p*-го обертона;

выявление интервалов времени, на которых огибающие колебания на ЧОТ и частоте *p*-го обертона превышают уровень 0,7 от своих максимальных значений;

определение временного интервала, на котором огибающие колебаний на ЧОТ и на частоте *p*-го обертона *совместно* превышают уровень 0,7;

"сшивание" фаз колебаний на ЧОТ и на частоте *p*-го обертона как функций времени в точках разрывов первого рода (скачки на величину 2π);

вычисление по формуле (4) разности фаз между колебаниями на ЧОТ и частоте *p*-го обертона;

построение графиков функций $<\Delta\Psi_{k1}^{p}(t)>$, $k=\overline{1, M}$, где угловые скобки символизируют определение среднего значений функций $\Delta\Psi_{k1}^{p}(t)$, $k=\overline{1, M}$, на выделенных с помощью анализа огибающих временных интервалах.

Экспериментальная часть

Цифровые записи РС четырех дикторов-мужчин (A,D,S,V) производились в специально оборудованном лабораторном помещении БГУИР, имеющем защиту от акустических шумов и шумов вибраций с уровнем <40 дБА.

Регистрация и анализ РС осуществлялись с помощью компьютера на базе процессора Pentium II.

Для ввода РС в использовались микрофон типа М-101 и 16-разрядная звуковая плата.

Частота дискретизации составляла 22050 Гц.

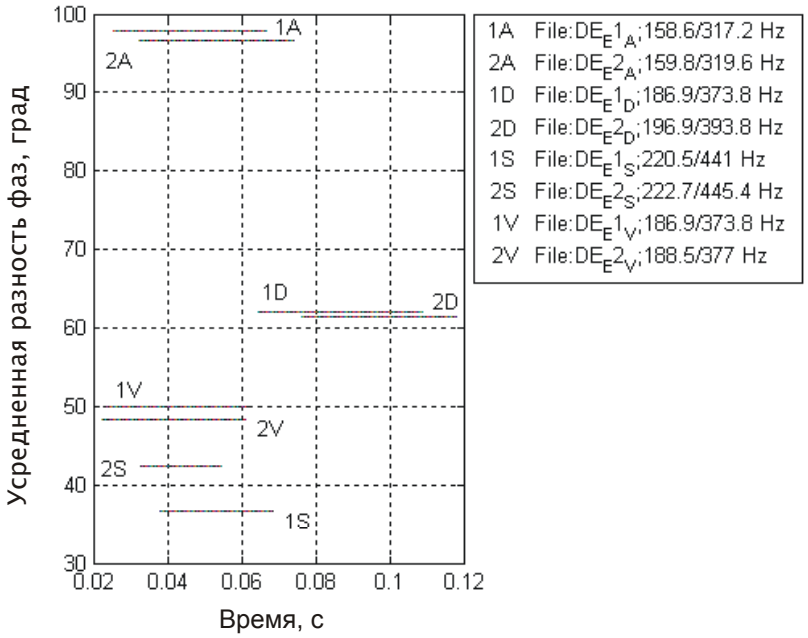
Для предварительного анализа временных и спектральных характеристик РС, а также временной селекции звуков и слов применялся пакет прикладных программ WaveLab 4.0d.

Спектральный, кепстральный анализ и межкомпонентная фазовая обработка цифровых записей выделяемых для исследования звуков и слов производились с помощью нескольких специально разработанных для этих целей рабочих программ.

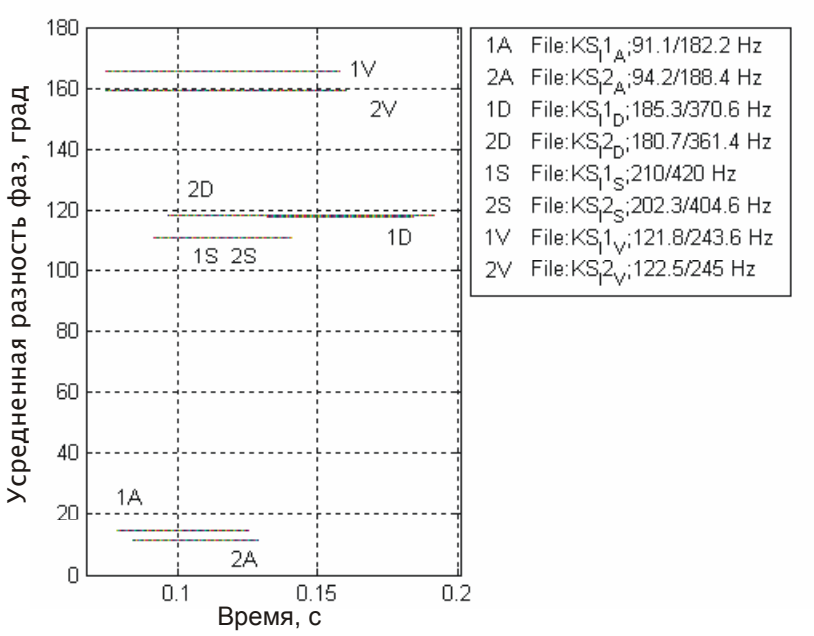
Результаты и их обсуждение

На рисунке в качестве примера результатов межкомпонентной фазовой обработки основного тона и обертонов приведены данные, полученные при анализе звуков "Э" и "И" в словах "ДЭ КСИ" (прочтение математического обозначения дифференциала $d\xi$).

С целью *наглядной* демонстрации *принципиальной* возможности использования информации, содержащейся в межкомпонентных фазовых соотношениях РС, для анализа выбрано по две реализации РС *с однообразным* (при восприятии на слух) *произнесением* упомянутых слов дикторами A, D, S и V.



а



б

Усредненные значения разности фаз между колебаниями на частоте F_0 основного тона и обертоном с частотой $2F_0$: *а*) для звука "э" в слове "дэ"; *б*) для звука "и" в слове "ξ" ("кси"). A, D, S, V — обозначения читающих дикторов

Как видно из графиков рисунка, в рассматриваемом случае наблюдается заметное различие усредненных разностей фаз между основным тоном и ближайшим к нему обертоном (p=2) как у различных дикторов при произнесении одного и того же звука, так и для различных звуков, произнесенных каждым диктором. Частоты основного тона и обертона для каждой реализации показаны в таблицах, помещенных справа от полей графиков.

Заключение

Проведенные исследования показали, что оценки разностей фаз между квазигармоническими составляющими РС на ЧОТ и кратных ей частотах оказываются незначительно изменяющимися на длительности тональных звуков, в общем случае различными для разных звуков, индивидуальными и устойчивыми для разных дикторов.

Количественные оценки степени устойчивости и информативности межкомпонентных фазовых характеристик требуют накопления статистически представительных данных на специально сформированных речевых базах. Вместе с тем имеющиеся у авторов рабочие материалы позволяют считать, что вариативность этих характеристик не явится препятствием для использования данных межкомпонентных фазовых измерений в РС при решении многих задач распознавания звуков речи и идентификации дикторов.

В случаях, когда традиционно используемые параметры речи (такие как ЧОТ, данные формантного анализа и др.) оказываются недостаточными для решения задач обнаружения, распознавания РС и идентификации дикторов, измерения и анализ указанных фазовых характеристик могут давать необходимую дополнительную информацию.

Важным качеством предложенных процедур обработки РС является их пониженная чувствительность (при отсутствии помех — полная нечувствительность) к изменениям интенсивности анализируемых РС.

PHASE RELATION BETWEEN FUNDAMENTAL TONES AND VOWEL SOUNDS OBERTONES

V.I. VARABYEU, G.U. DAVYDAU, YU.V. SHAMGIN

Abstract

It is shown by digital recordings processing of vowel sounds from different speakers, that difference of phases evaluation between fundamental tone and obertone carrying information, which can be used for listening discrimination and speaker identification.

Литература

1. *Сапожков М.А.* Речевой сигнал в кибернетике и связи (Преобразование речи применительно к задачам связи и кибернетики). М.: ГИЗЛ по вопросам связи и радио, 1963. 451 с.
2. *Воробьев В.И., Климов А.В.* //Радиотехника. 1986. № 2. С. 19–22.
3. *Воробьев В.И., Климов А.В.* // XXXIII Всесоюз. межвузовская науч.-техн. конф.: Тез. докл. Т. 1. Ч. 2. ТОВВМУ им. С.О. Макарова. Владивосток, 1990. С. 156.
4. *Воробьев В.И., Суданов П.М.* // XXXV Всесоюз. межвузовская науч.-техн. конф.: Тез. докл.. Т. 1. Ч. 1. ТОВВМУ им. С.О. Макарова. Владивосток, 1992. С. 50–51.
5. *Воробьев В.И.* / Материалы XIII научно-технического семинара РНТОРЭС им. А.С. Попова "Статистический синтез и анализ информационных систем". Рязань, Рязанская государственная радиотехническая академия, май 1994. С.75–78.
6. Анализ межкомпонентных фазовых соотношений в речевых сигналах: Отчет о НИР (ГБЦ 96-3023) / Белорусский государственный университет информатики и радиоэлектроники: БГУИР; Руководитель Воробьев В.И. Минск, 1996. 36 с. ГР № 19963577.
7. *Дворянкин С.* // Открытые системы, № 3, 2000.
8. *Дегтярев Н.П.* Параметрическое и информационное описание речевых сигналов. Минск, 2003. 216 с.