

# SPEAKER IDENTIFICATION BY CHARACTERISTICS OF LINEAR PREDICTION RESIDUAL

Fedorov A.V., Omelchenko A.V.  
Kharkov National University of Radio Electronics  
14, Lenin Ave., Kharkov, 61119, Ukraine  
Ph.: (057) 7021429, e-mail: father80@mail.ru

*Abstract* — Synthesis and research of text-independent adaptive algorithms to identify speakers by characteristics of linear prediction residual are described. As the informative features we use coefficients of Fourier transform of an autocorrelation function of a linear prediction residual signal processed with different time window functions. Experimental tests on the proposed identification algorithms are carried out.

## ИДЕНТИФИКАЦИЯ ДИКТОРОВ ПО ХАРАКТЕРИСТИКАМ ОСТАТКОВ ЛИНЕЙНОГО ПРЕДСКАЗАНИЯ

Фёдоров А. В., Омельченко А. В.  
Харьковский Национальный Университет Радиоэлектроники  
просп. Ленина, 14, Харьков, 61119, Украина  
тел.: (057) 7021429, e-mail: father80@mail.ru

*Аннотация* — Описывается синтез и исследование текстонезависимых адаптивных алгоритмов идентификации дикторов по характеристикам остатков линейного предсказания. В качестве информативных признаков используются коэффициенты преобразования Фурье автокорреляционной функции остатков линейного предсказания, пропущенных через различные временные окна. Выполняется экспериментальная проверка предложенных алгоритмов идентификации.

### I. Введение

Задаче идентификации личности по голосу уделяется значительное внимание [1, 2]. Этот интерес практически не ослабевает на протяжении последних 30 лет. Несмотря на это к настоящему моменту времени окончательного решения проблемы не найдено. Однако современный уровень развития инфотелекоммуникаций предполагает внедрение новых услуг, включающих в себя, управление своим счетом по телефону, удаленный доступ к информации и т.д. В связи с этим, актуальной задачей является разработка алгоритмов идентификации абонентов как мобильной, так и классической телефонной связи.

Целью настоящей работы является создание текстонезависимой системы идентификации дикторов по характеристикам остатков линейного предсказания. Поставленная задача решается в три этапа. На первом этапе формируется система информативных признаков. На втором — синтезируется правило принятия решения и на третьем — разрабатываются алгоритмы вычисления решающих статистик.

### II. Основная часть

В настоящее время для анализа речевых сигналов широко используется метод линейного предсказания (ЛП). Показав высокую эффективность в задачах распознавания речи [2], метод ЛП оказался приемлемым также и для автоматического опознавания дикторов [2]. В результате анализа речевого сигнала таким методом удастся получить комбинированную информацию о важных параметрах речевой активности конкретного человека, а именно: импульсы возбуждения голосовой щели, формантные частоты и полосы формантных областей [1]. Кроме того, методом ЛП могут быть оценены параметры основного тона голоса [2].

Пусть непрерывный сигнал  $s(t)$  дискретизируется с интервалом  $\Delta$ , т.е.  $s_n = s(n\Delta)$ , тогда значение речевого сигнала может быть представлено так [2]

$$s_n = \hat{s}_n + e_n, \quad \hat{s}_n = \sum_{k=1}^p a_k s_{n-k}, \quad (1)$$

где  $\hat{s}_n$  — предсказанное значение,  $a_k$  — коэффициенты предсказания,  $e_n$  — ошибка (остатки) предсказания,  $p$  — порядок модели ЛП. Выражение (1) и составляет суть метода ЛП [2].

В рамках модели линейного предсказания синтез речевого сигнала может быть выполнен с использованием рекурсивного линейного цифрового фильтра с передаточной функцией

$$H(z) = A (1 - \sum_{k=1}^p a_k z^{-k})^{-1} \quad (2)$$

где  $A$  — коэффициент усиления. Сигналом возбуждения  $u_n$  служит либо последовательность импульсов, формируемая генератором импульсов, либо шум, формируемый генератором белого шума. Эти два источника сигнала моделируют вокализованные и невокализованные фрагменты речевого сигнала. Ошибка предсказания  $e_n$  в рамках данной модели совпадает с функцией возбуждения.

Задача синтеза и исследования алгоритмов идентификации дикторов в рамках модели ЛП может быть сформулирована так. Реализации речи эталонных и идентифицируемых дикторов представлены сигналами, дискретизированными с частотой 8 кГц, 16 битов на отсчет. Остатки ЛП формируются решетчатый фильтром [2] 10-го порядка по сегментам длительностью 30 мс. Классификация речи на вокализованные и невокализованные сегменты выполняется автокорреляционным методом. Необходимо синтезировать адаптивные алгоритмы идентификации дикторов, устойчивые к возможным частотным искажениям исходного речевого сигнала в тракте передачи.

В задачах идентификации дикторов в качестве информативных признаков используются характеристики речевого сигнала, отражающие индивидуальность голоса. Согласно [1] основные физические проявления индивидуальности следует искать в спектральных и формантных, а также временных и амплитудных характеристиках сигнала.

Энергетический спектр  $|S(\omega)|^2$  звукового сигнала

$$|S(\omega)|^2 = |V(\omega)|^2 \cdot |T(\omega)|^2, \quad (3)$$

где  $|V(\omega)|^2$  — энергетический спектр сигнала голового возбуждения,  $T(\omega)$  — передаточная функция речевого тракта [3]. Таким образом, спектральные характеристики как остатков ЛП, так и исходного речевого сигнала содержат информацию об импульсах голосового возбуждения, форма и характер следования которых являются индивидуально дифференцирующими свойствами диктора [1].

Процедура формирования системы информативных признаков включает в себя следующие этапы: 1) формирование согласно (1) сигнала остатков ЛП; 2) интегрирование исходного речевого сигнала и остатков ЛП с целью компенсации эффекта дифференцирования осуществляемого губами говорящего. Передаточная функция цифрового интегрирующего фильтра имеет вид  $I(z) = (1 - \alpha z^{-1})^{-1}$ ,  $\alpha \approx 1$  [3]; 3) выделение информативных признаков из преобразованных сигналов исходной речи и остатков линейного предсказания, полученных на втором этапе. Кроме того, для исследования влияния частотных искажений речи, возникающих в тракте передачи, речевой сигнал на этапе 2 пропускнулся через цифровой дифференцирующий фильтр с передаточной функцией  $D(z) = 1 - \beta z^{-1}$ ,  $\beta = 0,9$ .

В настоящей работе в качестве информативных признаков использованы коэффициенты преобразования Фурье автокорреляционной функции преобразованных речевого сигнала и остатков ЛП, пропущенных через различные временные окна. Если через  $x_k$ ,  $k = 0, n-1$  обозначить координаты вектора информативных признаков  $\vec{x} = (x_0, x_1, \dots, x_{n-1})$  размерности  $n$ , то

$$x_k = \text{Re}(\sum_{m=0}^{n-1} r(m) w(m) e^{-2\pi k m / n}), \quad (4)$$

где  $r(m) = \sum_{l=0}^{n-1-m} s(l) s(l+m)$  — автокорреляционная функция сигнала  $s(l)$ , полученного на втором этапе формирования системы информативных признаков,  $w(m)$  — временное окно, форма которого определяется одним из следующих соотношений:  $w(m) = 1 - m/(n-1)$  (треугольное окно);  $w(m) = 1/2[1 + \cos(\pi m/[n-1])]$  (окно Хеннинга);  $w(m) = 0,54 + 0,46 \cos(\pi m/[n-1])$  (окно Хэмминга).

Правило принятия решения, полученное в рамках адаптивного подхода, имеет вид

$$j = \text{argmax}_{i=1, M} \max_a \{ \ln W(\vec{x} | i, a) \}, \quad (5)$$

где  $j \in \{1, 2, \dots, M\}$  — принятое решение об идентифицируемом дикторе,  $M$  — число дикторов,  $W(\vec{x} | i, a)$  — функция правдоподобия вектора информативных признаков  $\vec{x}$  с координатами определяемыми соотношением (4),  $a$  — параметр, учитывающий неопределенность уровня сигнала. Основываясь на решающем правиле (5) можно предложить ряд алгоритмов идентификации [5]:

$$A_1: j = \min_{i=1, M} (\ln \sum_{k=0}^{n-1} x_k / x_k^{(i)} + 1/n \sum_{k=0}^{n-1} \ln x_k^{(i)}); \quad (6)$$

$$A_2: j = \min_{i=1, M} \sum_{k=0}^{n-1} (\sqrt{x_k} - a \cdot \sqrt{x_k^{(i)}})^2. \quad (7)$$

Алгоритм  $A_1$  был получен в предположении, что элементы вектора  $\vec{x}$  являются независимыми показательно распределенными случайными величинами. Справедливость этого предположения подкрепляется тем фактом, что отсчеты спектрограмм стационарных временных рядов асимптотически независимы и имеют  $\chi_2^2$ -распределение [4]. В свою очередь алгоритм  $A_2$  легко получить считая элементы  $\vec{x}$  принадлежащими евклидовому пространству  $\mathbf{R}^n$  и выражая  $W(\vec{x} | i, a)$  через расстояние между точками этого пространства.

В настоящей работе выполнена экспериментальная проверка алгоритмов идентификации (6) и (7) для различных систем информативных признаков с учетом частотных искажений речи, возникающих в тракте передачи. В эксперименте использовалось 44 реализации 11 дикторов мужчин длительностью 18 с. каждая. 11 из 44 реализаций использовались на этапе обучения, остальные — на этапе идентификации. В таблице 1 приведены оптимизированные по временному окну данные о вероятности правильной идентификации дикторов, соответствующие указанным условиям проведения эксперимента.

Табл. 1.

Table 1.

Алгоритм Algorithm	Анализируемый сигнал Signal to analyze			
	Без частотных искажений речи в тракте передачи Without channel distortions		С частотными искажениями речи в тракте передачи Channel distortions are present	
	Речевой сигнал Speech signal	Сигнал остатков ЛП LP residual	Речевой сигнал Speech signal	Сигнал остатков ЛП LP residual
(6)	0,94	0,97	0,82	0,88
(7)	0,94	0,97	0,79	0,94

### III. Заключение

Анализ результатов экспериментальных исследований показывает, что точность идентификации дикторов по спектральным характеристикам остатков линейного предсказания выше точности идентификации дикторов по аналогичным характеристикам исходного речевого сигнала. Кроме того, использование остатков линейного предсказания позволяет компенсировать частотные искажения речевого сигнала, вносимые трактом передачи с неравномерной частотной характеристикой.

### IV. References

- [1] Ramishvili G.S. *Avtomaticheskoe opoznavanie govornashhego po golosu* [Automatic identification speaker by voice]. Moscow, Radio i svjaz', 1981. 224 p.
- [2] Markel J.D., Gray A.H. *Linear prediction of speech*. Berlin, New York, Springer-Verlag, 1976. 288 p.
- [3] Fant G. *Acoustic Theory of Speech Production*. Mouton, 1970. 311 p.
- [4] Brillinger D.R. *Time series. Data analysis and theory*. Holt, Rinehart and Winston, Inc., 1975. 487 p.
- [5] Fedorov A.V., Omelchenko A.V. Sintez i issledovanie algoritmov identifikacii diktorov po harakteristikam ostatkov linejnogo predskazanija [Synthesis and investigation of algorithms of identification of speakers by linear predictive residues characteristics]. *Radioelektronika i informatika*, 2006, No 4(35), pp. 71-75.