ՀԱՅԱՍՏԱՆԻ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԶԳԱՅԻՆ ԱԿԱԴԵՄԻԱ
НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК АРМЕНИИ
NATIONAL ACADEMY OF SCIENCES OF ARMENIA
ДОКЛАДЫ ԶԵԿՈՒՅՑՆԵՐ REPORTS

Հատոր Том Volume 107 2007 № 2

ИНФОРМАТИКА

УДК 621.395

Г. А. Бабаян

Комплексный метод оценки степени схожести сравниваемых сегментов речевого сигнала

(Представлено академиком А. Т. Кучукяном 23/I 2007)

Ключевые слова: *речевой сигнал, скремблер, временные перестановки, линейное предсказание, корреляция, минимальное расхождение*

С развитием тенденции увеличения доли оперативной информации, обусловленной бурным ростом личных связей в экономической деятельности людей, усиливается потребность в обеспечении конфиденциальности речевого обмена, ведущегося в основном по каналам тональной частоты.

Эффективным способом защиты речевых сообщений является использование устройств, обеспечивающих их защиту путем криптографических преобразований. К одним из таких устройств относятся аналоговые скремблеры, которые передают в канал связи преобразованный во времени и частотной области речевой сигнал, содержащий элементарные непрерывные сегменты исходного сигнала [1]. В них вся частотная полоса разбивается на несколько полос, и в каждой выделенной полосе производятся временные перестановки, чем достигается приемлемая стойкость защищаемой информации [1].

В работе [2] подробно рассмотрены все возможные варианты частотных преобразований сигнала и приведены способы автоматизации процесса восстановления фонограмм, подвергнутых таким преобразованиям. В [3] приведены некоторые способы восстановления фонограмм, подвергнутых временным перестановкам.

Целью данной статьи является развитие известных методов восстановления фонограмм, подвергнутых временным перестановкам, базирующимся на минимальном расхождении спектральных и волновых параметров сигнала

на смежных участках в точках разрывов непрерывного сигнала, а также оценка эффективности этих методов для нахождения места разрывов, называемых "точками коммутации", и непрерывных отрезков (сегментов), являющихся истинным продолжением предыдущего (смежный сегмент).

На рис. 1 схематично показан принцип временных перестановок в "окне" с длительностью T, в котором нумерованные сегменты открытого сообщения $X(t)$ переставлены в шифрованном $Y(t)$ сообщении [1].

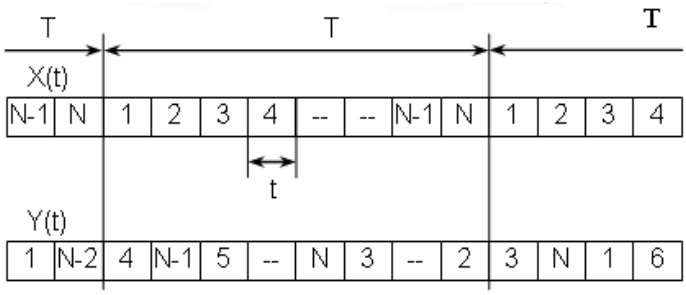


Рис.1. Пример перестановки "сегментов" t в "окне" с длительностью T

Задача восстановления фонограммы, подвергнутой временным перестановкам, сводится к поочередному нахождению смежных сегментов. Критерием поиска смежного сегмента может служить максимальное сходство начала смежного сегмента с окончанием предыдущего.

При автоматическом поиске смежного сегмента данный критерий можно определить следующим образом: смежным сегментом является сегмент, начальная часть которого имеет наименьшее расхождение волновых и спектральных параметров с параметрами окончания предыдущего сегмента из числа всех возможных претендентов (сегментов, расположенных в одном и том же "окне").

Критерий же нахождения "точек коммутации" при их автоматическом поиске определится следующим образом: если при сравнении двух соседних сегментов одинаковой длительности расхождение их волновых и спектральных параметров превосходит определенный уровень, то граница, разделяющая их, является "точкой коммутации".

Для эффективного использования второго критерия при автоматическом поиске "точек коммутации" необходимо вести попарное сравнение соседних сегментов, сканируя шифрованную фонограмму с шагом, меньшим длительности сегментов.

В работе [4] предложен метод идентификации речевых фрагментов, который может быть применен для поиска смежного сегмента с использованием подобного критерия, основанный на сравнении спектральных парамет-

ров с помощью обобщенного показателя, представляющего собой многоэлементную функцию с разными "весами" элементов в виде

$$P_{par} = \sum_{i=1}^{m} a_i d_i, \tag{1}$$

где P_{par} - параметр, характеризующий степень схожести сравниваемых сегментов; d_i, $i = \overline{1,m}$ элементы, используемые для сравнения спектральных срезов сегментов; a_i - весовой коэффициент i-го элемента; m - число элементов спектра.

В качестве параметров спектральных срезов, полученных с помощью БПФ, взяты:

- количество одноименных элементарных частотных полос спектральных срезов сравниваемых сегментов, расхождение амплитуд которых не превышает: 2db для первого, 4db для второго, 6db для третьего параметров в частотном диапазоне 250-700Гц;

- количество точно совпадающих по частоте уровневых максимумов (гармоник основного тона) в частотном диапазоне 250-700Гц;

- количество точно совпадающих по частоте уровневых минимумов в частотном диапазоне 250-700Гц;

- количество совпадающих по частоте уровневых максимумов в частотном диапазоне 250-700 с точностью одной элементарной частотной полосы;

- количество одноименных элементарных частотных полос спектральных срезов сравниваемых сегментов, расхождение амплитуд которых не превышает 2db, в частотном диапазоне 700-1250Гц.

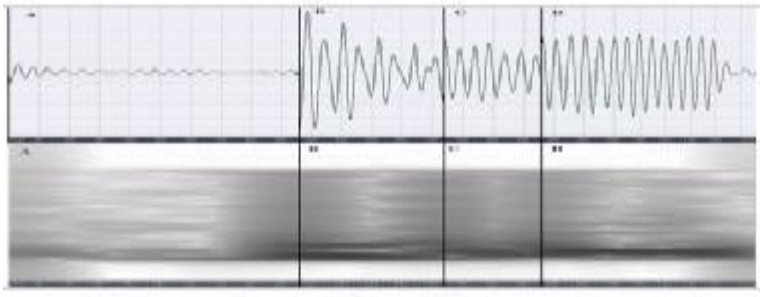
Коэффициенты каждого элемента a_i, $i = \overline{1,7}$ были получены эмпирически по методике, предложенной в работе [4], и равны: $a_1 = 0.576$; $a_2 = 0.652$; $a_3 = 0.674$; $a_4 = 0.478$; $a_5 = 0.554$; $a_6 = 0.532$; $a_7 = 0.367$.

Описанный метод, условно назовем его **параметрическим**, на практике, при сравнении 10мс участков искомых вокализованных смежных сегментов, показал в среднем 66.37% достоверность. Причины ошибок кроются в содержании сравниваемых сегментов. Укажем основные встречающиеся формы сигнала в сегменте шифрованной фонограммы, в случае которых **параметрический** метод может выдать ошибочный результат.

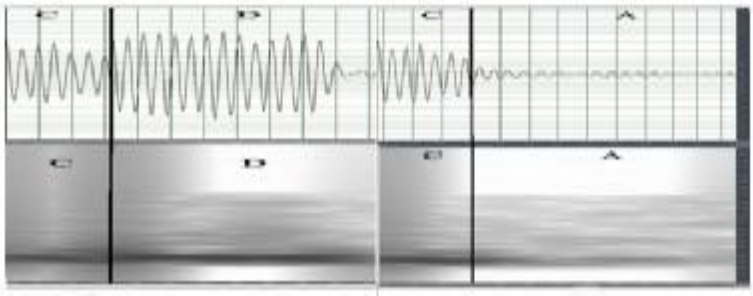
На рис. 2 приведены волновые и под ними соответствующие спектральные картины одной из характерных форм расположения сигнала в сравниваемых сегментах, когда **параметрический** метод приводит к ошибкам.

На рис. 2, а показан кусок шифрованной фонограммы, содержащей четыре переставленных сегмента. С использованием **параметрического** метода в качестве смежного сегмента для сегмента С был неверно найден сегмент D (рис. 2, б), тогда как истинным продолжением сегмента С является

сегмент A (рис. 2, в).



а) отрезок шифрованной фонограммы



б) неправильно в) правильно

Рис. 2 Пример ошибки, полученной при использовании **параметрического** метода

Первая причина ошибки связана с формой сигнала сегмента A, характерной для окончания вокализованного участка речи, когда амплитуда сигнала в нем резко убывает, тогда как в сегменте C, для которого A является истинным продолжением, она почти не меняется, и сегменты сильно отличаются усредненными амплитудами.

Вторая причина кроется в отсутствии на этих же участках явно выраженных гармоник основного тона (чередующиеся темные полосы на спектре). И поскольку в **параметрическом** методе как расхождение амплитуд сигнала, так и количество совпадаемых гармоник основного тона имеют достаточно большой вес, то обобщенный коэффициент сравнения для пары C, A будет мал по отношению к паре C, D, в которых оба указанных элемента спектра отличаются существенно меньше.

В качестве второго способа решения поставленной задачи, в настоящей статье предлагается использовать **корреляционный** метод, оценивающий степень сходства двух сигналов с помощью взаимно-корреляционной функции [5,6].

Чтобы избежать ошибки, связанной со сдвигом фаз, на практике корреляцию приходится находить, устанавливая несколько различных относительных задержек. При этом истинным считается наибольшее из полученных значений корреляции [5]. Поскольку речевой сигнал на сравниваемых

вокализованных участках длительностью 10 мс можно рассматривать как периодический [7], то для определения более точной корреляции целесообразно в одном из сравниваемых сегментов последовательность отсчетов циклически сдвигать влево, при этом крайний левый отсчет сигнала перемещать в правый конец. В этом случае корреляция будет определяться как

$$r_{12}(j) = \frac{1}{N} \sum_{n=0}^{N-1} x_1(n) \cdot x_2(n+j), \quad j = \overline{0, N-1}, \tag{2}$$

где $x_2(n+j)$ - сигнал $x_2(n)$, смещенный на j позиций влево, $r_{12}(j)$ - взаимная корреляция $x_1(n)$ и смещенного на j позиций сигнала $x_2(n+j)$. Из выражения (2) очевидно, что взаимная корреляция двух функций зависит от абсолютных значений сигнала, что может привести к очевидным ошибкам при сравнении сигналов с большой разницей амплитуд. Для исключения подобных ошибок необходимо нормировать взаимную корреляцию $r_{12}(j)$:

$$P_{12}(j) = \frac{r_{12}(j)}{\frac{1}{N} \cdot \sqrt{\left[\sum\limits_{n=0}^{N-1} x_1^2(n) \cdot \sum\limits_{n=0}^{N-1} x_2^2(n+j) \cdot\right]}}, \quad j = \overline{0, N-1}. \tag{3}$$

Далее корреляция расчитывается пошагово, сдвигая последовательность $x_2(n)$ на полный цикл, т.е. $j = \overline{0, N-1}$, и в качестве окончательного выбирается наибольшее из полученных значений корреляции:

$$P_{Cor} = \max_j \{P_{12}(j)\}, \quad j = \overline{0, N-1}. \tag{4}$$

Известно, что в речевом сигнале не бывает резких скачков по усредненному значению амплитуд [7]. Установив порог для амплитуды, ниже которой сигнал игнорируется, можно принять, что анализируемый сегмент является смежным, если удовлетворяются следующие условия:

$$\begin{cases} P_{Cor} = \max\limits_j \{P_{12}(j)\} > P_t, \quad j = \overline{0, N-1}, \\ A_{x_1} - A_{x_2} < A_t, \\ A_{x_1} > -40db, \end{cases} \tag{5}$$

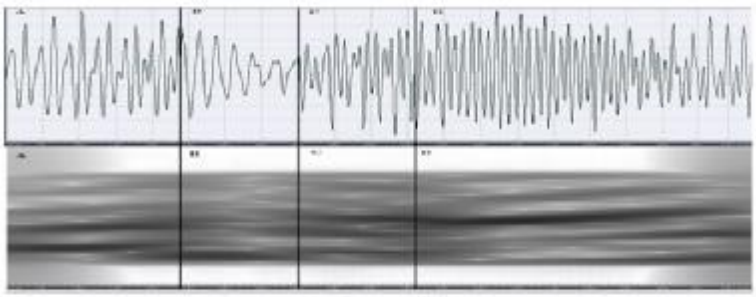
где P_t - минимальный порог значения корреляции, A_t - максимальный разброс среднеквадратичных значений амплитуд сравниваемых сегментов, A_{x_1}, A_{x_2} - среднеквадратичные значения амплитуд соответственно 1-го и 2-го сегментов. Значение P_t было приравнено наименьшей величине взаимной корреляции вычисленных значений из примерно 850 правильно найденных пар сегментов и составило $P_t = 0.57$.

Для нахождения значения A_t в местах нешифрованного сигнала, с резко меняющейся огибающей сигнала (в основном в конце или начале вокализованного сегмента), были рассчитаны разницы среднеквадратичных

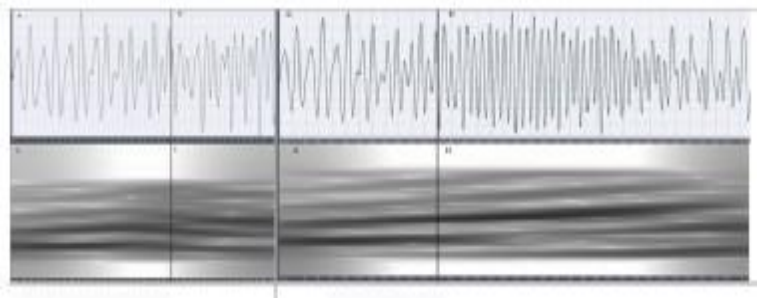
значений амплитуд 10 миллисекундных сегментов. Из примерно 5000 полученных значений A_t было выбрано самое большое $A_t = 4.18db$.

Для оценки данной методики проводились исследования на зашифрованных по ранее известному закону фонограммах. Количество правильно найденных смежных сегментов составило 66.8%. При этом иногда истинное продолжение находилось в тех случаях, когда другие методы выдавали ошибку. Например, в случае, рассмотренном на рис. 2, **корреляционным** методом, в отличие от **параметрического**, был правильно найден смежный сегмент A (рис.2, в).

Ошибочные же результаты с применением корреляционного метода были получены в сегментах, содержащих формы сигнала, приведенные на рис. 3.



а) отрезок шифрованной фонограммы



б) неправильно в) правильно

Рис. 3 Пример ошибки, полученной при использовании корреляцинного метода

На отрезке шифрованной фонограммы приведенной на рис. 3, а, продолжением для сегмента A **корреляционным** методом был ошибочно найден сегмент C (рис. 3, б), тогда как истинным продолжением сегмента A является сегмент D, который правильно был найден **параметрическим** методом. В этом примере причиной ошибки является относительно большая крутизна гармоник основного тона (это хорошо видно на спектральной картинке 4, в, справа).

Эффективность **корреляционного** метода уменьшается также при слабой выраженности старших гармоник основного тона, что существенно снижает

информативность волнового представления сигнала.

Вышеприведенные примеры подтверждают предположение о том, что **корреляционный** метод эффективно "работает" при волновом представлении сигнала и практически не учитывает его спектральную форму.

Приведем еще одну методику для решения поставленной задачи, основанную на известном методе кодирования **с линейным предсказанием** (Linear Predictive Coding LPC), суть которого состоит в том, что текущий отсчет речевого сигнала аппроксимируется линейной комбинацией предшествующих отсчетов. Коэффициенты линейного предсказания $\{a_k\}$ являются параметрами возбуждающего фильтра модели речеобразования [8], следовательно, они несут в себе информацию о характерных свойствах речевого сигнала.

Другое описание фильтра, приведеное в [9,10], основано на вычислении так называемых линейных спектральных параметров порядка P, представляющих собой упорядоченный набор из P чисел, принимающих значения из конечного интервала $[0.f/2]$, где f - частота отсчетов речевого сигнала. Спектр речевого сигнала при этом образуется перемножением линейчатого спектра возбуждающего сигнала и спектра, соответствующего голосовому тракту с формантными резонансными частотами, и, следовательно, также является линейчатым, а его огибающая характеризует передаточную функцию голосового тракта, включая формантные частоты. Такой подход позволил Итакуре и Сугамуре в 1979г. предложить метод линейных спектральных пар LSP (Line Specrtum Pair), которые определенным образом связаны с формантными частотами речевого сигнала [10].

Линейные спектральные параметры получаются из коэффициентов линейного предсказания разложением импульсной характеристики фильтра анализа коэффициентов линейного предсказания $A(Z)$ на сумму двух полиномов, определения которым даны в [9,10]. Из передаточных функций прямой и обратной фильтров модели речеобразования формируются следующие полиномы:

$$P(Z) = A(Z) - B(Z) = 1 + (a_M - a_1)Z^{-1} + \cdots + (a_1 - a_M)Z^{-M} - Z^{-(M+1)},$$

$$Q(Z) = A(Z) + B(Z) = 1 + (a_M + a_1)Z^{-1} + \cdots + (a_1 + a_M)Z^{-M} + Z^{-(M+1)}. \qquad (6)$$

Корни полиномов $P(z)$ и $Q(z)$ определяются путем поиска нулевых частот, т.е. частот, при которых эти полиномы обращаются в нуль. Поскольку $Z = e^{jw}$, то каждому корню в Z -плоскости соответствуют определенные положения в частотной плоскости ω_i, ν_i, между которыми сохряняется соотношение

$$0 = \omega_1 < \nu_1 < \omega_2 < \nu_2 < \cdots < \omega_i < \nu_i < \pi,$$

где ω_i - корни нечетного полинома $P(Z)$; ν_i - корни четного полинома $Q(Z)$. Частоты ω_i и ν_i образуют **спектральную пару**. Если две соседние частоты ω и ν расположены достаточно близко или даже равны, то эта спектральная пара будет соответствовать полюсу фильтра голосового тракта, т.е. местоположению форманты. Чем ближе расположены эти частоты, тем ширина полосы форманты становится уже, а амплитуда больше, что позволяет использовать их в качестве параметров при определении степени схожести двух сегментов по минимальному расхождению их средних значений [9].

Среднее значение расхождения ω_i

$$\omega_{jc} = \frac{1}{N}\sum_{i=1}^{N}|\omega_{0i} - \omega_{ji}|, \tag{7}$$

где ω_{jc} - среднее значение расхождения ω_i j-го сегмента в группе "претендентов" на смежный сегмент; ω_{0i} - корни полинома $P(Z)$ сегмента из той же группы, для которой осуществляется поиск смежного сегмента; ω_{ji} - корни полинома $P(Z)$ j-го сравниваемого сегмента группы; N - количество корней. Для речевого сигнала порядок линейного предсказания 10 считается оптимальным, и число корней при этом будет $N = 5$ [8,10]. Аналогично для ν_i

$$\nu_{jc} = \frac{1}{N}\sum_{i=1}^{N}|\nu_{0i} - \nu_{ji}|. \tag{8}$$

При анализе результатов поиска смежного сегмента в зашифрованных фонограммах с использованием линейных спектральных пар было выявлено, что кроме этих параметров также эффективно можно использовать как средние ω_j и ν_j:

$$\psi_j = \frac{\omega_j + \nu_j}{2}, \tag{9}$$

так и абсолютные значения их разности:

$$\sigma_j = |\omega_j - \nu_j|. \tag{10}$$

Поскольку линейные спектральные пары не несут в себе информацию об амплитудном значении сигнала, то с целью исключения сравнения сегментов с большой разницей амплитуд воспользуемся критерием допустимого абсолютного значения разницы среденеквадратичных значений амплитуд сравниваемых сегментов

$$\Delta A = \left|\frac{1}{n}\sqrt{\sum_{i=1}^{n}S_{1i}^2} - \frac{1}{n}\sqrt{\sum_{i=1}^{n}S_{2i}^2}\right|, \tag{11}$$

где S_{1i} и S_{2i}, соответственно, значения отсчетов первого и второго сравниваемых сегментов; n - количество отсчетов в сегментах.

Для учета всех вышеперечисленных параметров метода **линейного предсказания** при оценке схожести сегментов удобно пользоваться обобщающим параметром, включающим эти параметры в нормализованном виде с соответствующими весовыми коэффициентами:

$$P_{LP} = a_1 p_1 + a_2 p_2 + a_3 p_3 + a_4 p_4 + a_5 p_5, \tag{12}$$

где P_1, P_2, P_3, P_4, P_5 - нормализованные значения параметров:

$$P_1 = \frac{\omega_{jc}}{\omega_{jc\,\max}};\ P_2 = \frac{\nu_{jc}}{\nu_{jc\,\max}};\ P_3 = \frac{\psi_j}{\psi_{j\,\max}};\ P_4 = \frac{\sigma_j}{\sigma_{j\,\max}};\ P_5 = \frac{\Delta A}{\Delta A_{\max}}.$$

Значения нормализующих чисел $(\omega, \nu, \psi, \sigma, A)$ находятся экспериментально из шифрованных фонограмм, с заранее известными истинными смежными сегментами:

$$\omega_{j\,\max} = 218.223;\ \nu_{j\,\max} = 334.28;\ \psi_{j\,\max} = 113.23;\ \sigma_{j\,\max} = 257.71;\ \Delta A_{\max} = 0.0141.$$

В формуле (12) "веса" a_i, $i = \overline{1,5}$, были определены экспериментально, пропорционально вероятности правильного нахождения смежного сегмента для каждого параметра:

$$a_1 = 0.421; \quad a_2 = 0.463; \quad a_3 = 0.252; \quad a_4 = 0.547; \quad a_5 = 0.495.$$

Очевидно, что минимальные расхождения сравниваемых сегментов тем меньше, чем меньше значения P_{LP}.

Метод **линейного предсказания** имеет два основных недостатка: во-первых, он не учитывает временную форму сигнала; во-вторых, поскольку для нахождения коэффициентов линейного предсказания с порядком 10 необходимо брать отсчеты сегмента с длительностью не менее 20 миллисекунд, то для более коротких сегментов часть отсчетов будет взята из соседних сегментов,что может привести к ошибкам.

Эксперименты с использованием метода **линейного предсказания** проводились на 12 фонограммах, произнесенных на трех языках разными дикторами. Фонограммы были зашифрованы по ранее известному алгоритму, с применением аналогового скремблера, использующего только временные перестановки с минимальной длиной сегмента 10 миллисекунд.

Для сравнения на рис.4 приведены графики результатов применения к одним и тем же фонограммам методов **параметрического, корреляционного** и **линейного предсказания**, где вертикальная ось - вероятность P правильного нахождения смежных сегментов для данной фонограммы, горизонтальная ось - номер N экспериментальной фонограммы.

Поскольку каждый из рассмотренных методов имеет разную эффективность в зависимости от содержания сравниваемых сегментов, то, используя

эти методы в совокупности, в виде одного комплексного параметра, можно достичь большей достоверности нахождения смежного сегмента. С этой целью представим комплексный параметр в виде

$$K = a_{par}P'_{par} + a_{cor}P'_{cor} + a_{LP}(1 - P'_{LP}), \tag{13}$$

где a_{par}, a_{cor}, a_{LP}, соответственно, весовые коэффициенты **параметрического, корреляционного** и метода **линейного предсказания**; P'_{par}, P'_{cor}, P'_{LP} - нормализованные значения параметра каждого метода.

Степень схожести сравниваемых сегментов прямо пропорциональна значению K.

В качестве a_{par}, a_{cor}, a_{LP} берутся средние значения правильно найденных смежных сегментов каждого метода, которые равны $a_{par} = 0.6637$; $a_{cor} = 0.668$; $a_{LP} = 0.6244$. В формулу (13) подставляются нормализованные значения P, получаемые делением каждого параметра на максимальное значение параметра данного метода: $N_{par} = 35.53$; $N_{cor} = 0.9914$; $a_{LP} = 7.26$.

Результаты, полученные на тех же фонограммах с преименением комплексного метода, на рис.4 показаны в виде ломаной линии. Из графика видно относительное примущество этого метода.

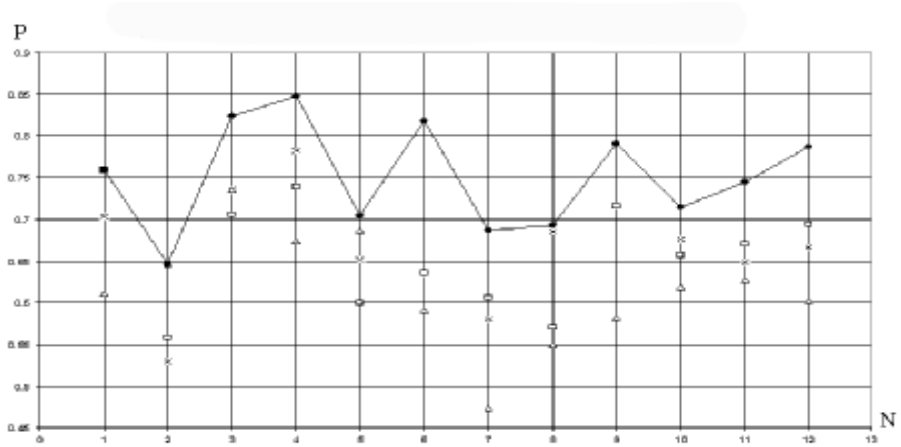


Рис. 4 Сравнительные результаты эффективности методов поиска смежного сегмента: × - параметрический, 2 - корреляционный, Δ - LPC, • - комплексный.

В заключение отметим, что предложенный комплексный метод оценки степени схожести сравниваемых сегментов речевого сигнала позволяет приблизительно с 75% вероятностью находить истинный смежный сегмент в автоматическом режиме восстановления фонограммы, шифрованной способом временных перестановок. Метод может найти практическое применение в комплексах инструментальных средств по автоматическому восстановлению фонограмм, шифрованных комбинированными аналоговыми скремблерами.

Ереванский научно-исследовательский институт математических машин
e-mail: bgrigor@ycrdi.am

Գ. Ա. Բաբայան

Ձայնային ազդանշանի համեմատվող կտորների նմանության գնահատման աստիճանի համալիր մեթոդ

Դիտարկված է ձայնային ազդանշանի համեմատվող կտորների նմանության գնահատման աստիճանի 3 մեթոդ՝ պարամետրական, կորելյացիոն և գծային կանխորոշման, որոնք հիմնված են ազդանշանի տարրապատկերային և ալիքային հատկությունների վրա: Ներկայացված է այդ մեթոդների արդյունավետության համեմատական գնահատական՝ 2 կտորների համեմատման հավաստիության արդյունքների տեսանկյունից: Առաջարկված է երեք մեթոդների արդյունքները միավորող գնահատման համալիր մեթոդ: Արդյունավետության ստուգման նպատակով բերված ժամանակային տեղափոխումների եղանակով գաղտնագրված իրական ֆոնոգրամայի օրինակի միջոցով ապացուցված է համալիր մեթոդի առավելությունը:

G. A. Babayan

Multimeter Method of Estimating the Similarity of the Compared Frames of Speech Signal

Three methods of estimating the degree of similarity of the compared frames of speech signal are examined: parametric, correlated and linear prediction which are based on a signal's peculiarities of spectral and wave representation. A comparative evaluation of its effectiveness is represented from the point of view of the results of the comparative truthfulness of two segments. A multimeter estimating method is proved through the example of checking its effectiveness for the restoration of real phonograms deciphered by means of time swapping.

Литература

1. Халяпин Д. Б. Защита информации. М. НОУ ШО "Баярд". 2004. 432 с.

2. Дарбинян А. А., Бабаян Г. А., Киракосян Т. А — Вестник Гос. инж. ун-та Армении. Серия "Моделирование, оптимизация, управление". 2006. Вып.9. Т.1. С.6-13.

3. Goldburg B., Dawson E., Sridharan S. In: Automated cryptanalysis of analog speech scramblers. Advances in Cryptology. Springer-Verlag. 1991. P. 422-430.

4. Бабаян Г. А., Дарбинян А. А., Киракосян Т. А., Оганесян А. Г. — Информационные технологии и управления. Ереван. 2005. №4-2. С 39-46.

5. Айфичер Э., Джервис Б. Цифровая обработка сигналов. 2-е изд. М. С-Пб. Киев. 2004. 992 с.

6. Сато Ю. Обработка сигналов. Первое знакомство. Изд. ДОДЭКА. 2002. 173с.

7. Фант Г. Акустическая теория речеобразования. М. "Наука". 1964. 284с.

8. Маркел Дж. Д., Грэй А. Х. Линейное предсказание речи. М. связь. 1980. 308с.

9. Коротаев Г. А. – Зарубежная радиоэлектроника. 1990. №3. С. 31-51.

10. Коротаев Г. А. – Зарубежная радиоэлектроника. 1991. №7. С. 3-31.