

**О. М. Ткаченко, к. т. н., доц.; О. Д. Феферман, студ.; С. В. Хрущак, студ.**

**А. Н. Ткаченко, к. т. н., доц.; О. Д. Феферман, студ.; С. В. Хрущак, студ.**

**A. Tkachenko, Cand. Sc. (Eng.), Assist. Prof.; O. Feferman, Student; S. Hruschak, Student**

## **ВОКОДЕР LSF ЗІ ШВИДКІСТЮ 1600 БІТ/С**

### **ВОКОДЕР LSF СО СКОРОСТЬЮ 1600 БІТ/С**

#### **LSF VOCODER OF SPEED 1600 BPS**

*Розглянуто вокодер, побудований на основі класичної моделі мовотворення за схемою лінійного завбачення з почерговим збудженням шумовими та періодичними сигналами. Модель вокодера реалізовано на персональному комп'ютері в реальному масштабі часу. Наведено основні функціональні характеристики вокодера та результати його практичних випробувань.*

*Рассматривается вокодер, построенный на основе классической модели речеобразования по схеме линейного предсказания с поочередным возбуждением шумовыми и периодическими сигналами. Модель вокодера реализована на персональном компьютере в реальном масштабе времени. Приведены основные функциональные характеристики вокодера и результаты его практических испытаний.*

*There had been considered the vocoder created on the basis of classical model of speech formation following the scheme of linear forecasting with an alternate noise and periodical signals excitation. The vocoder model is realized on line on PC. The paper also presents the vocoder basic functional characteristics and the results of its practical testing.*

#### **Вступ**

Сучасні досягнення в галузі створення високопродуктивних процесорів для обробки сигналів дозволяють суттєво вдосконалювати методи та алгоритми цифрової обробки мови з метою її ущільнення. Потреби в зменшенні швидкості передачі мовних сигналів виникають у військовій галузі та інших відомствах для забезпечення конфіденційності переговорів.

Для низькошвидкісних вокодерів (швидкість передачі до 2000 біт/с), які займають нішу службового зв'язку, характерна синтетична якість мови. Основним якісним показником роботи таких пристроїв є розбірливість синтезованої мови [1], яка має бути не нижча 75 %.

У вокодера, що розглядається, швидкість передачі параметрів 1600 біт/с. Вокодер побудовано на основі класичної моделі мовотворення, детально розглянутої в [2], за схемою лінійного завбачення з почерговим збудженням шумовими та періодичними сигналами. Модель вокодера працює на персональному комп'ютері в реальному масштабі часу, що дає змогу виконувати експериментальні дослідження роботи вокодера в локальній мережі, а також проводити випробування вокодера згідно з методикою [3].

#### **Вступление**

Современные достижения в области создания высокопродуктивных процессоров для обработки сигналов позволяют существенно усовершенствовать методы и алгоритмы цифровой обработки речи с целью ее сжатия. Необходимость уменьшения скорости передачи речевых сигналов возникает в военной области и других ведомствах для обеспечения конфиденциальности переговоров.

Для низкоскоростных вокодеров (скорость передачи до 2000 бит/с), которые занимают нишу служебной связи, характерно синтетическое качество речи. Основным качественным показателем работы таких устройств является разборчивость синтезированной речи [1], которая должна быть не ниже 75 %.

У представляемого вокодера скорость передачи параметров 1600 бит/с. Вокодер построен на основе классической модели речеобразования, которая детально рассмотрена в [2], по схеме линейного предсказания с поочередным возбуждением шумовыми и периодическими сигналами.

Модель вокодера працює на персональному комп'ютері в реальному масштабі часу, що дає можливість виконувати експериментальні дослідження в локальній мережі, а також проводити випробування вокодера згідно методики [3].

### Introduction

Modern achievements in the sphere of creation of highly efficient processors for signals processing allow substantial improvement of digital speech processing methods and algorithms with the purpose of its compression. The necessity of speed reduction of speech signals transmission arises in military and other spheres to ensure confidentiality support of private talks.

The synthetic speech quality is typical for the low-speed vocoders (with the transmission speed less than 2000 bps) which are used in official communications. The main qualitative factors of such devices is the synthesized speech intelligibility [1], which should be not less than 75 %.

The vocoder under consideration has data transmission speed of 1600 bps. It is created on the basis of classic model of speech production detailed in [2] using linear forecast with an alternate excitation by noise and periodical signals. Vocoder model runs on line on PC enabling to conduct the experimental researches in the intranet and carrying out vocoder tests in accordance with methods [3].

### Математична модель вокодера

Формально задачу компресії мови можна представити в двох варіантах. У першому варіанті необхідно синтезувати систему, що мінімізує швидкість передачі  $V$  із збереженням заданої якості  $A_0$ . У другому варіанті ставиться задача побудови системи, яка б забезпечила максимальну якість звукового сигналу  $A$  для заданої швидкості  $V$ . В обох варіантах вводяться додаткові обмеження, найважливішими з яких є:

- величина затримки сигналу  $\tau$  не повинна перевищувати максимально допустиму величину  $\tau_0$  (в телефонії – 200 мс);
- швидкодія реальних пристроїв  $N$ , що їх застосовують для компресії, не повинна перевищувати задану величину  $N_0$  (кілька десятків MIPS для сучасних DSP);
- обсяг пам'яті  $M$  для зберігання команд і даних не повинен перевищувати величину  $M_0$  (від десятків кілобайт до десятків мегабайт);
- стійкість до збоїв в каналах зв'язку  $S$  має бути не нижчою за задану величину  $S_0$ .

$$\begin{aligned} 1) & V \rightarrow \min; \quad A \geq A_0; \quad \tau \leq \tau_0; \quad N \leq N_0; \quad M \leq M_0; \quad S \geq S_0; \\ 2) & A \rightarrow \max; \quad V = \text{const}; \quad \tau \leq \tau_0; \quad N \leq N_0; \quad M \leq M_0; \quad S \geq S_0. \end{aligned} \quad (1)$$

Класична модель мовотворення ґрунтується на окремому описі голосового тракту і функції збудження [4]

$$s(n) = \sum_{k=1}^p a_k s(n-k) + G u(n), \quad (2)$$

де  $s(n)$  – відлік мовного сигналу (отсчет речевого сигнала, speech signal sample);  $u(n)$  – функція збудження (функція возбуждения, stimulation function);  $G$  – коефіцієнт підсилення (коэффициент усиления, gain factor);  $p$  – порядок моделі (порядок модели, model order).

Наведена модель містить лише полюси в своїй передавальній функції та не містить нулів. Але, як показано в [4], якщо порядок моделі  $p$  достатньо високий, вона дозволяє отримати достатньо точний опис для всіх звуків мови. Тому значення  $p$  обирають досить великим (8...14).

Задача ущільнення зводиться до оцінки параметрів (передавальної функції голосового тракту) та функції збудження (тон чи шум, якщо тон, то який період основного тону). Оскільки в процесі мовотворення параметри голосового тракту доволі швидко змінюються в часі, оцінка параметрів виконується кадрами. Їх тривалість вибирається такою, щоб протягом одного кадру звуковий сигнал можна було розглядати як стаціонарний випадковий процес. У сучасних вокодерах вона становить 10...40 мс.

Параметри моделі лінійного завбачення оцінюються автокореляційним методом з використанням рекурсії Левінсона—Дарбіна [4]. Отримані коефіцієнти лінійного завбачення (LPC)  $a_k$  є параметрами передатної функції  $A_p(z)$ , що описує голосовий тракт людини

$$A_p(z) = 1 + \sum_{k=1}^p a_k z^{-k}. \quad (3)$$

Хоча ці параметри можна після квантування безпосередньо передавати в канал зв'язку, в сучасних системах такий підхід практично не застосовується. Це пов'язано з тим, що LPC виявляються дуже нестійкими до каналних завад в системах зв'язку. Крім того, в процесі квантування чи інтерполяції цих параметрів можуть порушуватися умови стійкості системи. З іншого боку, LPC мають важливе теоретичне значення, оскільки, як буде показано далі, саме в просторі LPC визначено міри розрізнюваності. Тому практично доцільним видається підхід, коли отримані LPC параметри перетворюються в лінійні спектральні частоти (LSF).

Згідно з [5], LSF (за іншою термінологією – LSP, лінійні спектральні пари) є коренями поліномів

$$P(z) = P_{p+1}(z) = A_p(z) + z^{-(p+1)} A_p(z^{-1}); \quad (4)$$

$$Q(z) = Q_{p+1}(z) = A_p(z) - z^{-(p+1)} A_p(z^{-1}). \quad (5)$$

Корені поліномів  $P(z)$  і  $Q(z)$   $x_j$  та  $y_j$  лежать на одиничному колі, тому

$$\begin{aligned} x_j &= \cos \omega_{2j-1} \pm i \sin \omega_{2j-1}; \\ y_j &= \cos \theta_{2j} \pm i \sin \theta_{2j}; \\ 1 &\leq j \leq p/2. \end{aligned} \quad (6)$$

Набори  $\{\omega_{2j-1}\}$  та  $\{\theta_{2j}\}$  за визначенням є LSF. Властивість перемежованості

$$0 < \omega_1 < \theta_2 < \omega_3 < \dots < \omega_{p-1} < \theta_p < \pi \quad (7)$$

робить LSF менш чутливими до каналних помилок порівняно з LPC, а також значно спрощує перевірку системи на стійкість.

### Результати дослідження моделі вокодера

Досліджувалася робота вокодера за умов тривалості кадра 20 і 30 мс. Аналізуються та передаються в канал зв'язку такі параметри: обвідна спектра мовного сигналу, отримана на основі лінійного завбачення, коефіцієнт підсилення, що характеризує енергію сигналу, та період основного тону.

Для опису формантної обвідної спектра використовується модель лінійного передбачування 10 порядку. Аналіз здійснюється на кадрах довжиною в 160 (240) відліків (один відлік подається 16 бітами) з попереднім зважуванням вікном Хемінга.

Хоча існують методи, що дозволяють аналітично отримувати LSF з LPC [6], у вокодері, що розглядається, перетворення LPC в LSF виконується за чисельним методом [7], який дозволяє безпосередньо отримувати квантовані значення з кодової книги LSF. При цьому в канал зв'язку передається відповідний індекс з кожного рядка кодової книги. В скалярному квантуванні кожен з 10 коефіцієнтів LSF квантується незалежно з використанням власної кодової книги, розмірністю 3, 4, 4, 4, 4, 3, 3, 3, 3, 3 біта відповідно. Таким чином, загальний обсяг інформації про спектральну обвідну сигналу складає 34 біти.

Кодову книгу побудовано за методом центроїдів. Для оцінки відстані між векторами в кодовій книзі та пошуку найближчого квантованого значення пропонується використовувати міру розрізнюваності двох векторів в просторі LPC [4]

$$D(\hat{a}, a) = (\hat{a} - a) \left( N \frac{R}{\hat{a} R \hat{a}^t} \right) (\hat{a} - a)^t, \quad (8)$$

де  $D(\hat{a}, a)$  – відстань між сегментами мови з параметрами передбачування (расстояние между сегментами речи с параметрами предсказания, distance between speech segments with such parameters)

$a = (1, a(1), a(2), \dots, a(p))$ ,  $\hat{a} = (1, \hat{a}(1), \hat{a}(2), \dots, \hat{a}(p))$ ;  $t$  – знак транспонування (знак транспонирования, transposition sign);  $N$  – кількість відліків на сегмент (число отсчетов на сегмент, samples per frame);  $R$  – автокореляційна матриця мовного сигналу розміром (автокорреляционная матрица речевого сигнала размером, speech signal autocorrelation matrix with dimension of)  $(p+1) \times (p+1)$ .

Фактично міра (8) дозволяє визначити зважену відстань між вектором  $a$  параметрів мовного сигналу та будь-яким вектором  $a$  кодової книги. Застосування міри (8) дало змогу отримати кількісну оцінку якості ущільненого сигналу, яка добре узгоджується з результатами випробувань на розбірливість.

Оцінка періоду основного тону (ОТ) здійснювалася за методом зворотної фільтрації (SIFT-методом) [4]. Вибір даного методу обумовлений його високою точністю, а також тим, що більшу частину даних, необхідних для роботи алгоритму, було отримано на попередніх етапах обробки. Аналіз ОТ ведеться в діапазоні від 4 до 20 мс. Знайдені значення автокореляційної функції для зручності подальшої обробки проходять сортування за спаданням. Ті значення, що перевищують поріг у 0,3, утворюють коло претендентів, серед яких відбувається подальший вибір величини ОТ. При цьому враховується, що в деяких випадках (особливо для жіночих голосів) мають місце помилки, пов'язані з переходом на подвоєне значення періоду ОТ. Тому в разі прийняття рішень беруться до уваги як абсолютні, так і відносні значення претендентів в поточному кадрі, а також їх значення в попередньому та наступному кадрах. Остаточне рішення про величину ОТ приймається після квадратичної інтерполяції автокореляційної функції близько максимального значення. Знайдена величина ОТ упаковується в 7 бітів та передається в канал зв'язку.

Визначення енергії сигналу (коефіцієнта підсилення) в поточному кадрі відбувається під час обчислення LPC. Знайдене значення також квантується 7-бітною величиною.

Загальний обсяг інформації, що необхідно передати протягом кожного кадру, складає  $34 + 7 + 7 = 48$  біт. Таким чином, швидкість передачі  $V$  становить

$$V = 48 \text{ біт} / 0,02 \text{ с} = 2400 \text{ біт/с для кадру тривалістю 20 мс}$$

$$V = 48 \text{ біт} / 0,03 \text{ с} = 1600 \text{ біт/с для кадру тривалістю 30 мс.}$$

Для практичної апробації результатів досліджень було створено програмну модель вокодера в середовищі програмування Visual C++. Net. Модель працює в локальній мережі і підтримує такі режими роботи:

- без компресії (швидкість 128000 біт/с);
- ADPCM (швидкість 32000 біт/с);
- LPC (швидкість 4800 біт/с);
- LSF, скалярне квантування (швидкість 2400 біт/с).

Передбачено можливість зміни протоколу «на льоту» без переривання розмови, при чому кожен з абонентів може незалежно обирати для себе режим компресії. Конфіденційність розмови забезпечується за допомогою протоколу шифрування AES.

Для оцінки розбірливості сигналу після ущільнення було проведено випробування згідно з методикою [3]. Чотирьом слухачам пропонувалося заповнити по дві таблиці, що містили набори тестових слів, отриманих у результаті компресії за різними методами. Результати випробувань наведено в табл.

### Результати випробувань системи ущільнення мовних сигналів

Результаты испытания системы сжатия речевых сигналов  
The results of speech signals compression system test

Метод компресії (метод сжатия; compression method)	Довжина кадру, мс (длина кадру, мс, frame duration, ms)	Швидкість, біт/с (скорость, бит/с; speed, bits/s)	Середня відстань між векторами (среднее расстояние между векторами; average distance between vectors) $D$	Розбірливість (разборчивость; legibility), %
Без компресії (без сжатия; without compression)	—	128000	—	100 %
LPC	20	4800	—	87 %
LPC	30	3200	—	85,6 %
LSF	20	2400	0,3678	83,5 %
LSF	30	1600	0,3941	82 %

### Математическая модель вокодера

Формально задачу сжатия речи можно представить в двух вариантах (1). В первом варианте необходимо синтезировать систему, которая минимизирует скорость передачи  $V$  с сохранением заданного качества  $A$ . Во втором варианте ставится задача построения системы, которая обеспечила бы максимальное качество звукового сигнала  $A$  для скорости  $V$ . В обоих вариантах вводятся дополнительные ограничения, среди которых наиболее важными являются:

- величина задержки сигнала  $\tau$  не должна превышать максимально допустимую величину  $\tau_0$  (в телефонии – 200 мс);
- быстродействие реальных устройств  $N$ , которые применяются для сжатия, не должно превышать заданную величину  $N_0$  (несколько десятков MIPS для современных DSP);
- объемы памяти для сохранения команд и данных  $M$  не должны превышать величины  $M_0$  (от десятков килобайт до десятков мегабайт);
- устойчивость к сбоям в каналах связи  $S$  должна быть не ниже заданной величины  $S_0$ .

Классическая модель речеобразования основывается на раздельном описании голосового тракта и функции возбуждения (2) [4].

Приведенная модель (2) содержит только полюса в своей передаточной функции, но не содержит нулей. Однако, как показано в [4], если порядок модели  $p$  достаточно высокий, она позволяет получить достаточно точное описание для всех звуков речи. Поэтому  $p$  выбирают достаточно большим (8—14).

Задача сжатия сводится к оценке параметров (передаточной функции голосового тракта) и функции возбуждения (тон или шум, если тон, то какой период основного тона). Поскольку в процессе речеобразования параметры голосового тракта достаточно быстро изменяются во времени, оценка параметров выполняется по кадрам. Их продолжительность выбирается такой, чтобы на протяжении одного кадра звуковой сигнал можно было рассматривать как стационарный случайный процесс. В современных вокодерах она составляет 10...40 мс.

Параметры модели линейного предсказания оцениваются автокорреляционным методом с использованием рекурсии Левинсона—Дурбина [4]. Полученные коэффициенты линейного предсказания (LPC)  $a_k$  являются параметрами передаточной функции  $A_p(z)$ , которая описывает голосовой тракт человека (3).

Хотя эти параметры можно после квантования передавать непосредственно в канал связи, в современных системах такой подход практически не используется. Это связано с тем, что LPC являются весьма чувствительными к канальным помехам в системах связи. Кроме того, в процессе квантования или интерполяции этих параметров могут нарушаться условия устойчивости системы. С другой стороны, LPC имеют важное теоретическое значение, поскольку, как будет показано дальше, именно в пространстве LPC определены меры различимости. Поэтому практически целесообразным представляется подход, при котором полученные LPC-параметры преобразуются в линейные спектральные частоты (LSF).

Согласно [5], LSF (по другой терминологии – LSP, линейные спектральные пары) являются корнями полиномов (4) и (5).

Корни полиномов  $P(z)$  и  $Q(z)$   $x_j$  и  $y_j$  располагаются на единичной окружности, поэтому справедливо (6).

Наборы  $\{\omega_{2j-1}\}$  и  $\{\theta_{2j}\}$  по определению являются LSF. Свойство перемежаемости (7) делает LSF менее чувствительными к канальным ошибкам по сравнению с LPC, а также значительно упрощает проверку системы на устойчивость.

### Результаты исследования модели вокодера

Исследовалась работа вокодера при условии продолжительности кадра 20 и 30 мс. Анализируются и передаются в канал связи такие параметры: огибающая спектра речевого сигнала, полученная на основе линейного предсказания, коэффициент усиления, который характеризует энергию сигнала, и период основного тона.

Для описания формантной огибающей спектра используется модель линейного предсказания 10 порядка. Анализ производится на кадрах длиной в 160 (240) отсчетов (один отсчет представляется 16 битами) с предварительным взвешиванием окном Хемминга.

Хотя существуют методы, которые позволяют аналитически получать LSF из LPC [6], в рассматриваемом вокодере преобразование LPC в LSF выполняется с помощью численного метода [7], который позволяет непосредственно получать квантованные значения из кодовой книги LSF. При этом в канал связи передается соответствующий индекс из каждой строчки кодовой книги. При скалярном квантовании каждый из 10 коэффициентов LSF квантуется независимо, с использованием

собственной кодовой книги, размерностью 3, 4, 4, 4, 4, 3, 3, 3, 3, 3 бита соответственно. Таким образом, общий объем информации о спектральной огибающей сигнала составляет 34 бита.

Кодовая книга построена по методу центроидов. Для оценки расстояния между векторами в кодовой книге и поиска ближайшего квантованного значения предлагается использовать меру различимости двух векторов в пространстве LPC (8) [4].

Фактически мера (8) позволяет определить взвешенное расстояние между вектором  $\hat{a}$  параметров речевого сигнала и любым вектором  $a$  кодовой книги. Использование меры (8) дало возможность получить количественную оценку качества сжатого сигнала, которая хорошо согласуется с результатами испытаний на разборчивость.

Оценка периода основного тона (ОТ) осуществлялась по методу обратной фильтрации (SIFT-метод) [4]. Выбор данного метода обусловлен его высокой точностью, а также тем, что большая часть данных, необходимых для работы алгоритма, была получена на предыдущих этапах обработки. Анализ ОТ ведется в диапазоне от 4 до 20 мс. Найденные значения автокорреляционной функции для удобства дальнейшей обработки проходят сортировку по убыванию. Те значения, которые превышают порог в 0,3, составляют множество претендентов, среди которых происходит дальнейший выбор величин ОТ. При этом учитывается, что в некоторых случаях (особенно для женских голосов) имеют место ошибки, связанные с переходом на удвоенное значение периода ОТ. Поэтому при принятии решений рассматриваются как абсолютные, так и относительные значения претендентов в текущем кадре, а также их значения в предыдущем и последующем кадрах. Окончательное решение о величине ОТ принимается после квадратичной интерполяции автокорреляционной функции около максимального значения. Найденная величина ОТ упаковывается в 7 битов и передается в канал связи.

Определение энергии сигнала (коэффициента усиления) в текущем кадре происходит во время вычисления LPC. Найденное значение также квантуется 7-битной величиной.

Общий объем информации, который необходимо передать на протяжении каждого кадра, составляет  $34 + 7 + 7 = 48$  бит. Таким образом, скорость передачи  $V$  составляет:

$$V = 48 \text{ бит} / 0,02 \text{ с} = 2400 \text{ бит/с для кадра продолжительностью 20 мс и}$$

$$V = 48 \text{ бит} / 0,03 \text{ с} = 1600 \text{ бит/с для кадра продолжительностью 30 мс.}$$

Для практической апробации результатов исследования была создана программная модель вокодера в среде программирования Visual C++. NET. Модель работает в локальной сети и поддерживает такие режимы работы:

- без компрессии (скорость 128000 бит / с);
- ADPCM (скорость 32000 бит/с);
- LPC (скорость 4800 бит/с);
- LSF, скалярное квантование (скорость 2400 бит/с).

Предусмотрена возможность изменения протокола «на лету» без прерывания разговора, причем каждый из абонентов может независимо выбирать для себя режим компрессии. Конфиденциальность разговора обеспечивается при помощи протокола шифрования AES.

Для оценки разборчивости сигнала после сжатия были проведены испытания согласно методике [3]. Четирем слушателям предлагалось заполнить по две таблицы, которые содержали наборы тестовых слов, полученных в результате сжатия с использованием разных методов. Результаты испытаний приведены в таблице.

### Vocoder mathematical model

The speech compression problem can be formally represented in two ways (1). In the first case it is necessary to synthesize the system, which minimizes the transmission speed  $V$  and at the same time preserves the given speech quality  $A$ . In the second case the task is to build the system ensuring the maximum quality of speech signal  $A$  at the given speed  $V$ . In both cases the additional limitations are introduced. The most important are the following:

- signal delay value  $\tau$  should not exceed the maximum allowed value  $\tau_0$  (in the telephony – 200 ms);
- processing speed  $N$  of real devices used for compression, must not exceed the given value  $N_0$  (several tens MIPS for the modern DSP);
- memory capacity  $M$  for data and commands storage must not exceed  $M_0$  (from the tens of kilobytes to the several tens of megabytes);
- stability to failures in the communication links  $S$  must not exceed the given value  $S_0$ .

The classic model of speech formation is based on the separate description of vocal track and the stimulation function (2) [4].

The given model (2) contains only poles in its transfer function but does not contain zeroes.

However, as is shown in [4], if model degree  $p$  is rather great, it allows to get the precise enough description for all sounds of speech. Therefore  $p$  should be great enough (8...14).

Compression task comes to the parameters estimation (voice track transfer function) and excitation function (periodic signal or noise, if periodic – than what is the pitch period). As the voice track parameters change rather quickly during the speed production, parameters estimation is accomplished in the frame diapason. Their duration should chosen in a way, which allows to consider the speech signal as stationary random process during one frame. In the modern vocoders it makes up 10–40 ms.

Parameters of linear forecast model are estimated with the help of autocorrelation method using Levinson—Durbin recursion [4]. The received linear prediction factors (LPC)  $a_k$  are transfer function parameters  $A_p(z)$ , describing human voice track (3).

Though these parameters can be transmitted directly to the communication channel after quantization, this approach isn't of common use in modern systems. That is caused by the fact that the LPC have great sensibility to the channel noise in communication systems. Besides, system stability may be broken during quantization or interpolation of these parameters. On the other hand, LPC have an important theoretic significance, because, as is shown bellow, the distinguishability measures are defined exactly in the LPC area. Therefore the approach of transforming the obtained LPC parameters to linear spectral frequencies (LSF) is preferable.

According to [5], LSF (the other name – LSP, linear spectral pairs) are the roots of polynomials (4) and (5).

The roots of polynomials  $P(z)$  and  $Q(z)$   $x_j$  and  $y_j$  are located on unit circumference, that's why (6) comes true.

The sets  $\{\omega_{2j-1}\}$  and  $\{\theta_{2j}\}$  according to their definition are LSF. The intersect property (7) makes LSF less sensible to the channel errors in comparison with the LPC, and also considerably simplifies the system stability check.

### The research results of vocoder model

There had been analyzed the vocoder functioning under conditions of frame duration of 20 and 30 ms. The following parameters had been analyzed and transferred to the communication channel: speech signal spectral envelope, received on the basis of linear prediction, amplification factor, which describes the energy of signal and its pitch.

The linear forecast model of tenth order is used for the description of the spectral envelope. Analysis is carried out on the frames with duration of 160 (240) samples (each sample is represented by 16 bits) with preliminary Hamming window weighting.

Though there are methods allowing an analytical receive of LSF from LPC [6], the LPC to LSF transformation is performed by numerical method [7] in the described vocoder allowing for direct obtaining of quantized values from LSF. In this case an appropriate index from the each row of codebook is transmitted to the communication channel. When using the scalar quantization, each of ten LSF factors is quantized independently, using its own codebook of such dimensions: 3, 4, 4, 4, 3, 3, 3, 3, 3 bits correspondingly. In this case the total information content describing the spectral envelope totals 34 bits.

The codebook is composed according to the method of centroids. There had been suggested the distinguishability measure of two vectors in the LPC area [4] (8) for the evaluation of the distance between vectors in codebook and search for the nearest quantized value.

The virtual measure (8) allows to calculate the balanced distance between speech signal vector  $\hat{a}$  and any other vector  $a$  of the codebook. The usage of measure (8) had given the possibility of receiving quantitative evaluation of compressed signal quality conforming to legibility test results.

Pitch estimation had been conducted using the inverse filtering method (SIFT-method) [4]. The choice of this method is determined by its high accuracy and also by the fact, that the greater part of data necessary for algorithm functioning was received during previous stages of process. Pitch analysis is accomplished within the range 4...20 ms. The obtained autocorrelation function values undergo the descending classification for the convenience of the further process. The values exceeding the threshold of 0.3, form the a great number of candidates, among which the subsequent pitch value selection is continued. At the same time it is necessary to consider that in some cases (especially for women voices) there are errors due to the detection of doubled pitch value. Therefore during the decision taking there had been considered absolute and relative numbers of candidates in the current frame and also their values in the previous and the next frame. Final decision on the pitch value is made after the square interpolation of autocorrelation function near it's maximum value. The obtained pitch value is packed into 7 bits and transferred to the communication channel.

The determination of signal energy (amplification factor) in the current frame occurs during the LPC computing. The obtained value is also quantized by 7-bit value.

The total transferred information content totals  $34 + 7 + 7 = 48$  bits. So, the transmission speed  $V$  is

$V = 48 \text{ bits}/0,02 \text{ s} = 2400 \text{ bps}$  for frame of duration of 20 ms and

$V = 48 \text{ bits}/0,03 \text{ s} = 1600 \text{ bps}$  for frame of duration of 30 ms.

There had been created the vocoder software model for the practical approbation of the research results using Visual C++. NET. This model works in the local area network and supports such operating modes:

- without compression (with the speed of 128000 bits per second);
- ADPCM (with the speed of 32000 bits per second)
- LPC (with the speed of 4800 bits per second)
- LSF, scalar quantization (with the speed of 2400 bits per second).

There had been foreseen the possibility of changing protocol "immediately" without the termination of conversation. It should be noted that each side can independently choose compression mode for itself. Conversation confidentiality is provided by the AES cryptography algorithm.

There had been conducted testing according to the method [3] for the evaluation of signal legibility after compression. The four listeners were suggested to fill in the two tables with test word sets containing words received after compression using different methods. Test results are shown in table.

### Висновки

Побудовано модель вокодера, що дозволяє здійснювати ущільнення мовних сигналів до швидкості 1600 біт/с. У ході експериментальних досліджень вивчалася робота даної моделі в локальній мережі для різних протоколів компресії та проводилися випробування моделі згідно з прийнятою методикою. Випробування показали достатню якість ущільненого мовного сигналу.

**Перспективи.** З огляду на стрімкий розвиток цифрової телефонії, найближчим часом слід очікувати появи вокодерів зі швидкістю передачі 1200 біт/с і 600 біт/с (фонемне розпізнавання). Розробка таких вокодерів потребує врахування тонких механізмів мовотворення, зокрема фонетичних особливостей конкретної мови. Необхідною умовою здійснення таких розробок є створення статистично репрезентативної бази даних мовних фрагментів української мови. Поява такої бази дасть потужний поштовх для розвитку досліджень у цьому напрямку в Україні.

### Выводы

Построена модель вокодера, позволяющая производить сжатие речевых сигналов до скорости 1600 бит/с. В ходе экспериментальных исследований изучалась работа данной модели в локальной сети для разных протоколов сжатия и проводились испытания модели согласно принятой методики. Испытания показали достаточное качество сжатого речевого сигнала.

**Перспективы.** Учитывая чрезвычайно быстрое развитие цифровой телефонии, в ближайшее время следует ожидать появления вокодеров со скоростью передачи 1200 бит/с и 600 бит/с (фонемное распознавание). Разработка таких вокодеров потребует учета тонких механизмов речеобразования, в частности фонетических особенностей конкретного языка. Необходимым условием осуществления таких разработок является создание статистически репрезентативной базы данных речевых фрагментов украинского языка. Появление такой базы даст мощный толчок для развития исследований в этом направлении в Украине.

### Conclusions

There had been created the model of vocoder allowing for speech signals compressing to the speed of 1600 bps. The operation of the given model in the local net for different protocols had been tested following the accepted methods during the researches period. The tests had shown high quality of compressed speech signal.

**Prospects.** Taking into consideration the great digital telephony development allows to assume, that the vocoders with transmission speed of 1200 bps and 600 bps (phonemic recognition) will appear in the nearest future. Development of such vocoders will take into consideration the complicated mechanisms of speech production, particularly phonetic peculiarities of specific language. The creation of statistically representative database of Ukrainian language speech fragments is the necessary condition of such development. The creation of such database will give the powerful incentive to the development of such researches in Ukraine.



## СПИСОК ЛІТЕРАТУРИ

## СПИСОК ЛИТЕРАТУРЫ

## REFERENCES

1. Сапожков М. А., Михайлов В. Г. Вокодерная связь. — М.: Радио и связь, 1983. — 248 с.
2. Deller J., Proakis J., Hansen J. Discrete-Time Processing of Speech Signals. — IEEE press, 2000. — 466 p.
3. ГОСТ 16600-72. Передача речи по трактам радиотелефонной связи. Требования к разборчивости речи и методы артикуляционных измерений.
4. Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов: Пер. с англ. — М.: Радио и связь, 1981. — 496 с.
5. Иванов В. Н. Вычисление линейных спектральных частот // Электросвязь. — 1997. — № 6. — С. 25—27.
6. Павлов О. И. Упрощение реализации метода линейных спектральных пар (частот) в линейном предсказании речи // Труды 3-й Международной конференции «Цифровая обработка сигналов и ее применение». — М., 2000. — С. 126—129.
7. Kabal P., Ramachandran R. P. The Computation of Line Spectral Frequencies Using Chebyshev Polynomials // IEEE Transactions on Acoustics, Speech, and Signal Processing. — 1986. — V. ASSP-34, № 6. — P. 134—137.

Рекомендована кафедрою обчислювальної техніки

Надійшла до редакції 16.06.05  
Рекомендована до друку 5.07.05

**Ткаченко Олександр Миколайович** — доцент кафедри обчислювальної техніки; **Феферман Олег Дмитрович, Хрущак Сергій Вікторович** — студенти Інституту інформаційних технологій та комп'ютерної інженерії.

Вінницький національний технічний університет.

**Ткаченко Александр Николаевич** — доцент кафедры вычислительной техники; **Феферман Олег Дмитриевич, Хрущак Сергей Викторович** — студенты Института информационных технологий и компьютерной инженерии.

Винницкий национальный технический университет.

**Olexandr Tkachenko** — Assistant Professor of the Chair of Computer Engineering; **Oleg Feferman, Serhiy Khruschak** — Students of Institute for Information Technologies and Computer Engineering.

Vinnitsia National Technical University