

УДК 534.782.001:621.39

А.Н. Продеус, канд. техн. наук

Частотное распределение формант украинской и русской речи

Произведено оценивание и сопоставление плотностей распределения формант по частоте для украинской и русской речи.

Evaluation and comparison of probability density functions for frequency formants distribution for Ukrainian and Russian speech is produced.

Ключевые слова: плотность распределения, форманты, украинская речь, русская речь, речевой сигнал.

Введение

Одной из разновидностей акустической экспертизы является оценка качества помещений и линий связи с использованием ряда критериев. При этом наиболее важным критерием качества является разборчивость (понятность) речи [1]. В качестве инструментального (объективного) метода оценки разборчивости речи часто используют формантный метод [2, 3], согласно которому артикуляционную (формантную) разборчивость речи оценивают при помощи соотношения:

$$A = \sum_{k=1}^K p_k \cdot w(E_k), \quad (1)$$

где p_k – вероятность пребывания формант в k -й полосе частот:

$$p_k = F(f_{vk}) - F(f_{hk}), \quad (2)$$

$F(f)$ – функция распределения вероятностей формант по частоте; $w(E_k)$ – коэффициент восприятия речи; E_k – разность эффективных уровней речевого сигнала и маскирующего его шума (отношение сигнал-шум) в k -й частотной полосе.

Как следует из соотношений (1)–(2), для оценивания артикуляционной разборчивости речи необходимо располагать сведениями о функциях $w(E_k)$ и $F(f)$, характеризующих вероятностные свойства формант. Для русской речи такая информация имеется [2–6]. Впрочем, неоднозначность этой информации столь существенна, что приходится говорить о существовании, по меньшей мере, трех версий формантного метода [1, 7–9]. Еще хуже обстоит дело с оцениванием разборчивости украинской речи, для которой вид функций $w(E_k)$ и $F(f)$ до недавнего времени вообще был неизвестен. Начало восполнения указанного пробела было

положено в работе [10], где впервые получены оценки функций распределения формант по уровню для украинской речи, а также произведено сопоставление оценок функций распределения формант по уровню для русской и украинской речи.

Целью данной работы является реализация следующего этапа исследований, а именно: оценка и сопоставление плотностей распределения формант по частоте для украинской и русской речи.

1. Методика исследований

Методика сбора данных для экспериментальных исследований описана в работе [10]. Напомним вкратце ее особенности. Запись речевых сигналов производилась для двух групп дикторов: трех мужчин и четырех женщин, практически одинаково хорошо владеющих русским и украинским языками. Зачитывались фонетически сбалансированные украинский и русский тексты, состоящие из фрагментов художественных произведений, технических и юридических публикаций. Русский и украинский тексты отличались по содержанию (т. е. русский текст не был переводом украинского текста, и наоборот). Дикторы читали текст по возможности слитно, сводя к минимуму паузы между словами и предложениями. Параметры записи речевых сигналов следующие: длительность реализации речевого сигнала 17 мин; частота дискретизации 22050 Гц; глубина квантования 16 бит.

Обработка записанных речевых сигналов осуществлялась поэтапно: 1) стационаризация речевого сигнала (для подавления непривильного изменения уровня сигнала во времени из-за усталости диктора или по иным причинам); 2) формантный анализ речевого сигнала; 3) оценивание плотности распределения $p(f) = dF(f)/df$ по результатам формантного анализа; 4) сопоставление и анализ полученных оценок функции $p(f)$.

Формантный анализ речевого сигнала – процедура весьма сложная, поэтому, по возможности подробно, рассмотрим этот этап исследований.

Сопоставление опубликованных результатов формантного анализа речевых сигналов

Графики оценок плотности распределения формант по частоте для русской речи, приве-

денные в ряде советских [2–4] и российских [5–6] литературных источников, показаны на рис. 1. Заметим, что советские исследователи оценивали частоту появления формант в полосах частот шириной 100 Гц, поэтому для рис. 1, а [2] следует полагать $p(f) = \Delta A_{100}(f) \cdot 0,01$, для рис. 1, в [4] $p(f) = A_e(f) \cdot 0,01$, и для рис. 1, г [5] $p(f) = A(f) \cdot 0,01$.

Показанный на рис. 1, б график оценки $p(f) = dF(f)/df$, где

$$F(f) = \begin{cases} 2,57 \cdot 10^{-8} \cdot f^{2,4}, & 100 < f \leq 400; \\ 1 - 1,074 \cdot \exp(-10^{-4} \cdot f^{1,18}), & 400 < f \leq 10000 \end{cases}$$

используется современными российскими исследователями [5–6] и представляет собой «сглаженную» версию графика, приведенного на рис. 1, а.

Нетрудно видеть, что графики рис. 1, в, г заметно отличаются от графиков рис. 1, а, б. Пояснить это можно различием как методических, так и аппаратурных погрешностей измерений, которые организовывались различными авторскими коллективами свыше пятидесяти лет тому назад. По причине давности этих исследований, а также в силу ряда иных причин, сегодня практически невозможно добыть сведения об особенностях методик этих измерений. Имеющиеся в открытой литературе сведения крайне скучны. Например, Н.Б. Покровский в работе [2] ограничивается следующим описанием: «Разбивают весь диапазон частот на

полоски, например, по 100 Гц, и подсчитывают относительное число формант (%) в каждой полоске. Результат такого подсчета даст кривую $\Delta A(f)$ » (рис. 1, а). Не ясна лишь одна «малость»: каким образом совершается такой подсчет. Не более откровенны по этому поводу и авторы работ [3–5].

Таким образом, приходится констатировать существенную неоднозначность информации о виде функции $p(f)$ для русской речи.

К сожалению, для украинской речи дело обстоит еще хуже – в известной нам литературе отсутствует информация о функциях $p(f)$.

Формантный анализ с использованием программы Praat

Одним из наиболее удобных и, как следствие, популярных программных инструментов для обработки речевых сигналов является компьютерная программа Praat, позволяющая, в частности, осуществлять так называемый «формантный анализ» [11].

На рис. 2, а показана форма речевого сигнала для фразы «електромагнітні коливання» (мужской голос), а на рис. 2, б приведены результаты формантного анализа. При этом были заданы следующие параметры алгоритма формантного анализа:

1) нижняя частота корректирующего фильтра $F_H = 50$ Гц (после этой частоты АЧХ фильтра возрастает со скоростью 6 дБ на октаву);

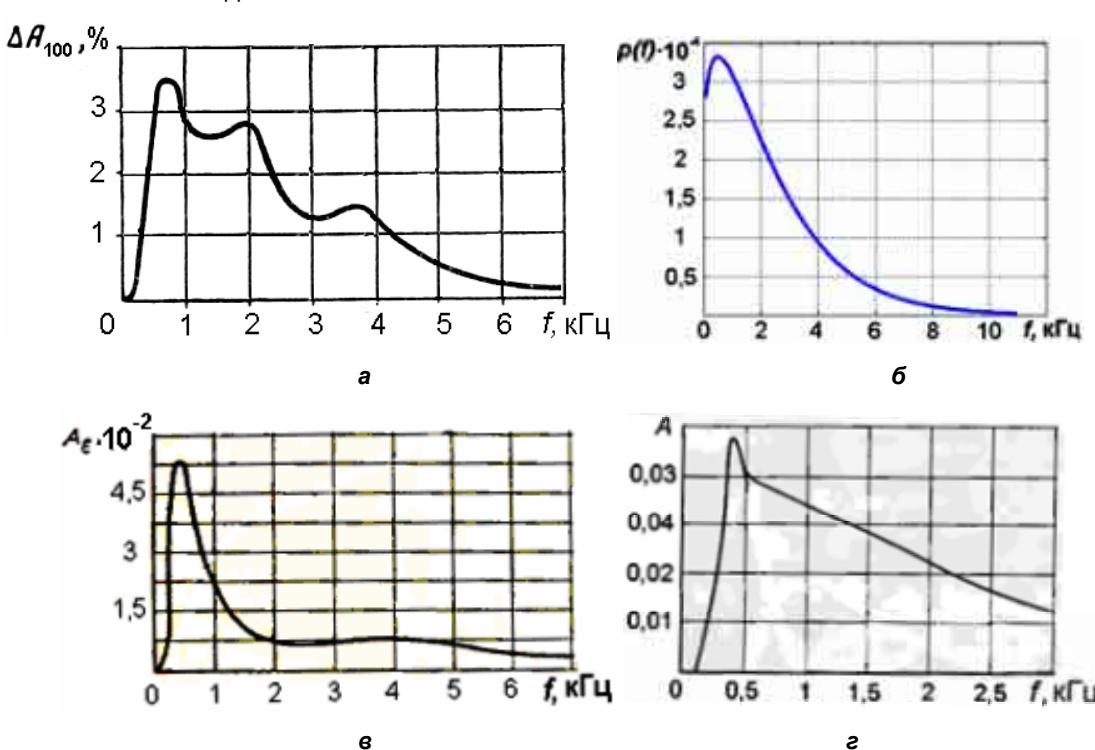


Рис. 1. Оценки плотности распределения формант по частоте

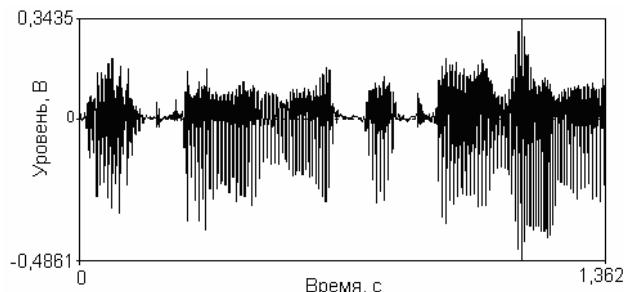
2) весовое окно данных, протяженностью 50 мс, имеет форму гауссовой кривой с эффективной шириной 25 мс;

3) сдвиг фреймов данных составляет 25 % протяженности весового окна данных;

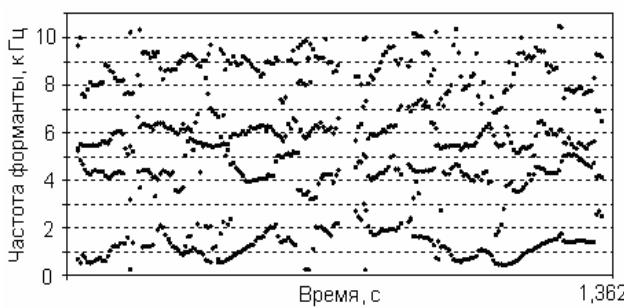
4) предполагается, что количество формант не превышает пяти;

5) максимальная частота рассматриваемого диапазона частот – 11025 Гц;

6) используется параметрическая оценка спектра Берга.



а



б

Рис. 2. Форма речевого сигнала (а) и его формантный портрет (б)

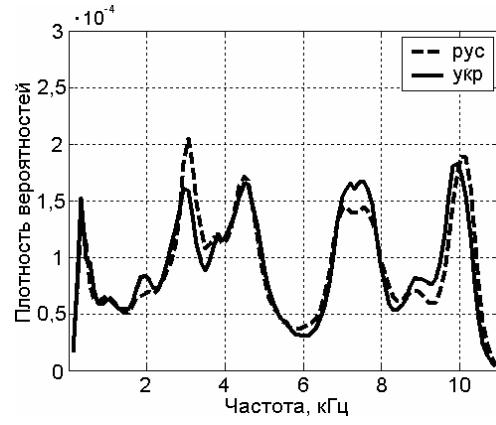
Нетрудно видеть, что результаты формантного анализа представляют собой набор траекторий формант на плоскости время-частота. Эти результаты можно сохранить и в виде текстового файла. На рис. 3 показан фрагмент такого файла, соответствующий речевому сигналу в пределах одного окна данных. Стоящее в первой строке число – это интенсивность сигнала, содержащегося в окне данных, число «4» означает количество обнаруженных (в пределах окна данных) формант, далее идут четыре пары чисел – частота максимума и ширина полосы частот каждой из обнаруженных формант.

Очевидно, что для нас информативными являются лишь частоты максимумов обнаруженных формант. На рис. 4 показаны полученные в результате обработки такого текстового файла (программа обработки приведена в Приложении 1) оценки плотности распределения формант по частоте $p(f)$ для диктора-женщины

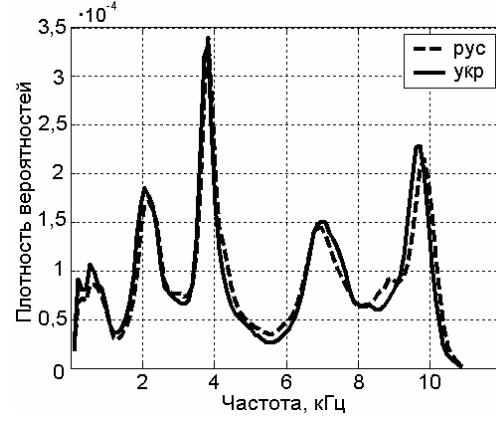
и диктора-мужчины, для русского и украинского языков.

0.0037507029548550495
4
532.4031016754192
842.3328327489753
4848.196226670095
1786.2350457109023
5535.429302896622
379.01373276614015
9972.973751665411
1090.3719745320195

Рис. 3. Фрагмент текстового файла с результатами формантного анализа



а



б

Рис. 4. Оценки функции $p(f)$ для диктора-женщины (а) и диктора-мужчины (б)

На основании полученных результатов можно предварительно заключить, что различие пола дикторов значительно сильнее сказывается на форме функций $p(f)$, нежели фонетическое различие русского и украинского языков.

О предыскажении речевого сигнала

Сравнивая графики рис. 4 и 1, нетрудно заметить, что наиболее принципиальным является отсутствие на рис. 4 глобального (и существенно выраженного) максимума функции $p(f)$ в окрестности 400...500 Гц.

Покажем, что причиной этого явления является процедура предыскажения речевого сигнала, производимая в программе Praat с помощью корректирующего фильтра, описываемого уравнением:

$$y_i = x_i - \alpha x_{i-1}, \quad \alpha = \exp(-2\pi F_h \Delta t), \quad (3)$$

где x_i – выборки корректируемого (исходного) речевого сигнала; y_i – выборки откорректированного речевого сигнала; F_h – частота, начиная с которой амплитудно-частотная характеристика (АЧХ) фильтра растет со скоростью 6 дБ на октаву (по умолчанию принято $F_h = 50$ Гц); Δt – период дискретизации. График АЧХ корректирующего фильтра приведен на рис. 5, где хорошо видно, что АЧХ фильтра действительно возрастает со скоростью 6 дБ на октаву в диапазоне частот 50...8000 Гц.

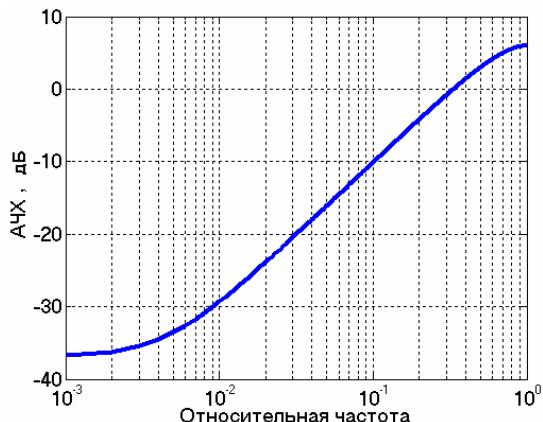


Рис. 5. АЧХ корректирующего фильтра

Между тем, анализ формы долговременного спектра речи B_p (рис. 6) позволяет сделать вывод, что в интервале 100...250 Гц уровень спектра растет со скоростью примерно 12 дБ на октаву, а в интервале 250...500 Гц изменение уровня спектра не превышает 3...4 дБ (здесь и далее термин «спектр речи» использован вместо эквивалентного, но громоздкого термина «спектральная плотность мощности речевого сигнала»). Лишь начиная с 500 Гц, уровень спектра убывает со скоростью 6 дБ на октаву. Таким образом, процедура предыскажения речевого сигнала обеспечивает равномерность спектра речи для частот выше 500 Гц. Как следствие – для частот выше 500 Гц обеспечивается возможность выявления даже слабых формант-

ных всплесков, ранее (до предыскажения) не наблюдавшихся из-за убывания уровня долговременного спектра речи.

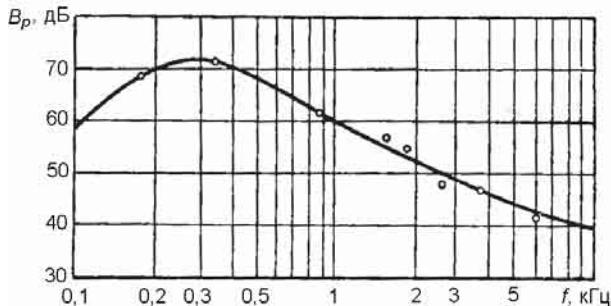


Рис. 6. Долговременный спектр речи [2]

Вместе с тем, очевидно, что предыскажение речевого сигнала приводит к дополнительному росту (со скоростью 6 дБ на октаву) крутизны спектра речевого сигнала в диапазоне 100...250 Гц. При этом форма спектральных пиков, находящихся в указанном диапазоне, искажается столь существенно, что многие спектральные пики перестают наблюдаться.

Если учесть, что диапазон 100...250 Гц соответствует частотам основного тона [2–4], приходим к выводу, что процедура предыскажения позволяет решить две задачи: 1) понизить частоту обнаружения низкочастотных спектральных пиков, соответствующих частотам основного тона; 2) «выровнять» вероятности обнаружения формант на частотах выше 500 Гц.

Следует, однако, помнить, что в диапазоне 250...500 Гц, где изменение уровня долговременного спектра речи не превышает 3...4 дБ, заключена первая форманта многих звуков речевого сигнала [2–4]. Очевидно, процедура предыскажения, искажая форму долговременного спектра в этом диапазоне, вносит систематическую погрешность в подсчет частоты появления этой форманты.

Чтобы ослабить этот эффект, можно поступить двояко: либо принять параметр F_h равным не 50 Гц (рис. 4), а 500 или даже 1000 Гц, либо вообще отказаться от процедуры предыскажения. Сопоставляя рис. 7, где показаны результаты осуществления обоих вариантов, с рис. 4, а, где для того же диктора показана ситуация предыскажения со значением $F_h = 50$ Гц, нетрудно прийти к выводу, что процедура предыскажения, выполненная в том или ином варианте, мало сказывается на значениях функции $p(f)$ для частот выше 500 Гц. А вот в частотном диапазоне 100...500 Гц процедура предыскажения может существенно (в данном конкретном случае – до пяти раз) изменить значения функции $p(f)$.

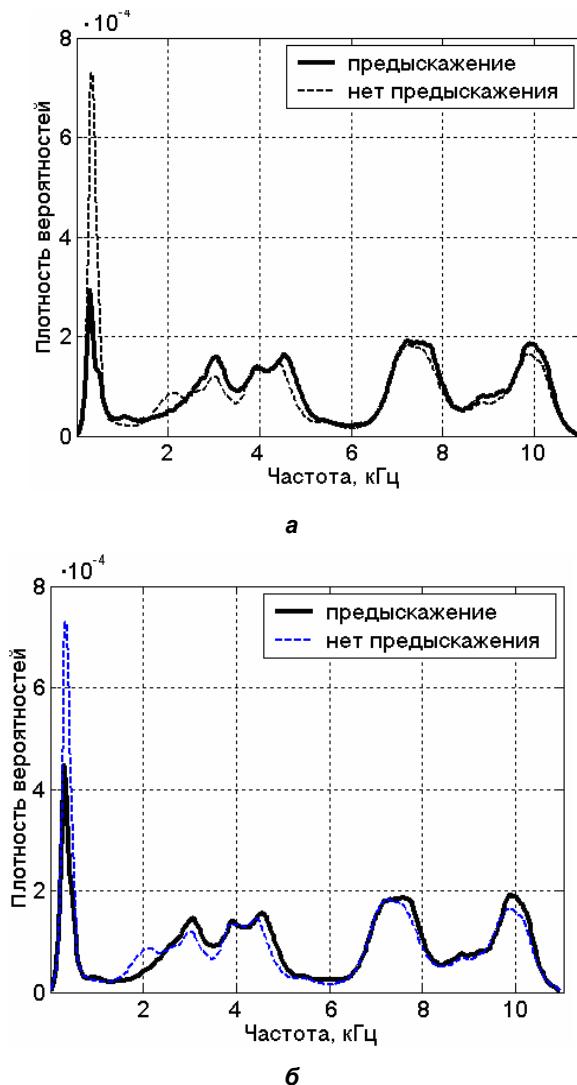


Рис. 7. Оценки функции $p(f)$ для $F_H = 500$ Гц (а) и $F_H = 1000$ Гц (б)

Формантный анализ с использованием среды Matlab

Поскольку в программе Praat управление процедурой предыскажения речевого сигнала при формантном анализе не предусмотрено, автором данной работы была разработана собственная компьютерная программа для среды Matlab (Приложение 2), с помощью которой и были получены представленные на рис. 7 результаты. Краткие характеристики этой программы:

1) используется параметрическая оценка спектра Берга, порядок которой принят равным 10, что позволяет обнаруживать до пяти формант в спектре речевого сигнала;

2) сегменты данных, взвешенных окном Блэкмана, состоят из 1024 выборок – при этом эффективная ширина окна близка 25 мс (при частоте дискретизации 22050 Гц); степень

перекрытия сегментов принята равной 75 %.

Представленные далее результаты исследований получены с использованием этой программы, без предыскажения речевых сигналов.

2. Результаты исследований

На рис. 8 представлены результаты сопоставления функций $p(f)$ для следующих случаев:

- 1) мужчины: русский, украинский языки (рис. 8, а);
- 2) женщины: русский, украинский языки (рис. 8, б);
- 3) мужчины и женщины: русский, украинский языки (рис. 8, в);
- 4) русский язык: мужчины, женщины (рис. 8, г);
- 5) украинский язык: мужчины, женщины (рис. 8, д).

Для количественной оценки степени различия сопоставляемых графиков, вычислялась относительная среднеквадратическая ошибка (ОСКО), значения которой приведены в заголовках каждого из рисунков. Сопоставляя форму приведенных графиков, а также значения соответствующих ОСКО, нетрудно заметить, что оценки функций $p(f)$ для украинской и русской речи весьма близки (ОСКО=0,2...0,3). Оценки функций $p(f)$ для мужчин и женщин также являются близкими, однако различие между ними заметно больше (ОСКО=0,5...0,7).

Выводы

При оценивании плотности вероятностей $p(f)$ распределения формант по частоте следует учитывать, что применение процедуры предыскажения речевого сигнала негативно сказывается на точности измерений функции $p(f)$ в области первой форманты.

Результаты проведенных экспериментальных исследований свидетельствуют о весьма высокой степени сходства вероятностных свойств русских и украинских речевых сигналов: оценки функций $p(f)$ для украинской и русской речи оказались весьма близкими (ОСКО=0,2...0,3); оценки функций $p(f)$ для мужчин и женщин также являются близкими, однако различие между ними заметно больше (ОСКО=0,5...0,7).

Представленные результаты носят предварительный характер, поскольку в экспериментальных исследованиях участвовало небольшое количество дикторов.

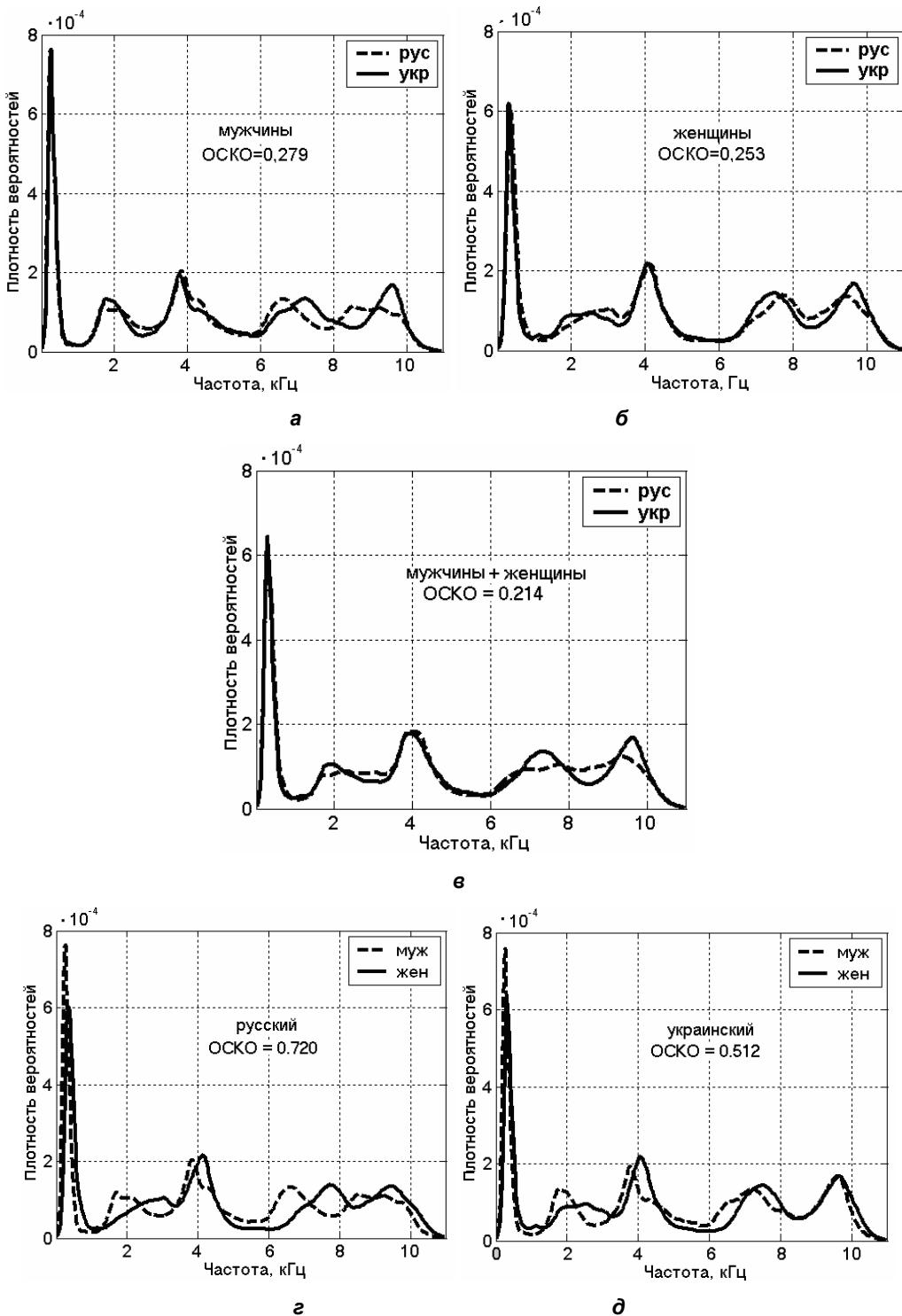


Рис. 8. Сопоставление оценок плотностей распределения формант по частоте

Литература

1. Дидковский В.С., Дидковская М.В., Продеус.А.Н. Акустическая экспертиза каналов речевой коммуникации. Монография. – К: Им-экс-ЛТД, 2008. – 418 с.
2. Покровский Н.Б. Расчет и измерение разборчивости речи. – М.: Связьиздат, 1962. – 390 с.
3. Сапожков М.А., Михайлов В.Г. Вокодерная связь. – М.: Радио и связь, 1983. – 247 с.
4. Быков Ю.С. Теория разборчивости речи и повышение эффективности радиотелефонной связи. – М.-Л.: Госэнергоиздат, 1959. – 351 с.
5. Железняк В.К., Макаров Ю.К., Хорев А.А. Некоторые методические подходы к оценке эффективности защиты речевой информа-

- ции // Специальная техника. – М.: 2000, № 4. – http://st.ess.ru/publications/4_2000/geleznk/gel_eznyak.pdf.
6. Каргашин В.Л. Совершенствование методических принципов оценки защищенности помещений от утечки речевой информации // Специальная техника. – 2001. – № 6. – http://www.bnti.ru/showart.asp?aid=579&lvl=0_3.03.04.
 7. Продеус А.Н., Гавриленко А.В., Дидковский В.С. Сравнительный анализ некоторых методов оценки разборчивости речи // Сб. трудов Акустического симпозиума "Консонанс-2007" – 2007. – С. 273–278.
 8. Продеус А.Н., Гавриленко А.В., Дидковский В.С. Сопоставление версий формантного метода оценки разборчивости речи // Электроника и связь. Тематический выпуск «Проблемы электроники». – 2008. – Ч.1. – С. 227–231.
 9. Продеус А.Н., Гавриленко А.В., Дидковский В.С., Самойленко Н.А. Формантный метод оценки разборчивости речи // Электроника и связь – 2007. – № 5. – С. 63–70.
 10. Дидковский В.С., Продеус А.Н. Сопоставление формантных свойств украинской и русской речи // Электроника и связь. Тематический выпуск «Электроника и нанотехнологии». – 2009. – Ч. 2. – С. 88–94.
 11. Praat: doing phonetics by computer. – <http://www.fon.hum.uva.nl/praat/>.

Национальный технический университет Украины
«Киевский политехнический институт»

Поступила в редакцию 22 мая 2009 г.

Приложение 1

Программа оценки плотности распределения формант по частоте (Praat)

```
% === formant_raspred_Praat.m === распределение формант по частоте ===
% 1) исходные данные - массив x - формантный портрет от программы Praat (burg);
% 2) сжатие массива x - выбрасывание всего, кроме формантных частот;
% 3) использование стандартной программы hist() для построения гистограммы;
% 4) нормирование гистограммы для получения оценки плотности вероятностей;
% ===== начало расчетов =====
i=1; z=[];
while i<length(x)
count = x(i);
y=x(i:i+2*count+1);
y(1+2*count+1)=[];
for j=1+2*count:-2:1
y(j)=[];
end
z=[z;y];
i=i+2*count+2;
end
% === нормирование графика и вывод на печать ===
[n,xout]=hist(z,100);
sumn=sum(n);
del=(max(xout)-min(xout))/100;
nf=n/sumn/del;
figure; plot(xout,nf); % график
```

Приложение 2

Программа оценки плотности распределения формант по частоте

```
% === formant_raspred_we.m === распред. формант по частоте ===
% 1) вычисление огибающей спектра мощности;
% 2) вычисление координат максимумов огибающей спектра мощности;
% 3) запись единиц для тех частот, в которые попали максимумы;
% 4) организация «окошек» по оси частот и подсчет кол-ва единиц в них;
% 5) нормирование полученной гистограммы по «площади»;
% 6) оценка вероятностей пребывания формант в октавных полосах частот;
% === прочтение речевого файла ===
```

```
filename = 'Boichuk_rus_stat'; % имя стационаризированного файла
x = wavread(filename); % чтение файла данных
z=wavread('chimes'); % чтение файла сигнализации (об окончании расчетов)
Fs = 22050; % частота дискретизации
% === посегментная обработка речевого файла ===
p = 10; nfft = 1024; % порядок полинома; параметр БПФ
nseg = 1024; % кол-во выборок сегмента
P1 = zeros(1,511);
window = blackman(nseg);
step = floor(nseg/4);
for j = 1:floor(length(x)/step)-3;
    xseg = x((j-1)*step+1:(j-1)*step+nseg); % сегмент сигнала
    xseg = xseg.*window;
    [Pxx,f] = pburg(xseg,p,nfft,Fs); % огибающая спектра мощности (Берг)
    dPxx = diff(Pxx); sdPxx = sign(dPxx); P1m1 = -diff(sdPxx)/2; % единица - максимум
    for i = 1:length(P1m1)
        if P1m1(i)<0
            P1m1(i)=0;
        end
    end
    P1 = P1 + P1m1';
end
P2=P1(2:length(P1)-1);
f2=f(2:length(f)-3);
f2=f2';
% === нормирование ===
delf = f2(2)-f2(1);
sumP2 = sum(P2);
P2=P2/sumP2/delf;
% === сглаживание по 5 элементам ===
P3 = filter([0.2 0.2 0.2 0.2 0.2],1,P2);
figure;
plot(f2,P3); % график гистограммы
grid on;
title(['Бойчук, жен., рус.'], 'FontSize', 14, 'FontName', 'MS Sans Serif');
xlabel(['Частота, Гц'], 'FontSize', 14, 'FontName', 'MS Sans Serif');
% === запись на диск распределения формант ===
formant_my_Boichuk_rus = struct('arg',f2,'func',P3);
save formant_my_Boichuk_rus formant_my_Boichuk_rus;
% === подсчет вероятностей в октавных полосах ===
n_gr = [4 9 17 34 66 131 260]; % номера спектр.выборок - вергн.границ окт.полос
v_gr = [8 16 33 65 130 260 length(P2)]; % номера спектр.выборок - вергн.границ окт.полос
p = zeros(1,7);
for k = 1:7
    for l = n_gr(k):v_gr(k)
        p(1,k) = p(1,k)+P2(l);
    end
    p(1,k)=p(1,k)*delf;
end
figure; bar(p); % график гистограммы
title(['Бойчук, жен., рус.'], 'FontSize', 14, 'FontName', 'MS Sans Serif');
xlabel(['Номер полосы частот'], 'FontSize', 14, 'FontName', 'MS Sans Serif');
% === запись на диск распределения вероятностей формант в октавных полосах ===
formant_pi_Boichuk_rus = struct('func',p);
save formant_pi_Boichuk_rus formant_pi_Boichuk_rus;
```