

Формантний аналіз та синтез МОВНИХ СИГНАЛІВ

Моделі голосового тракту

Визначення формант

Параметри формант

Способи візуалізації формант

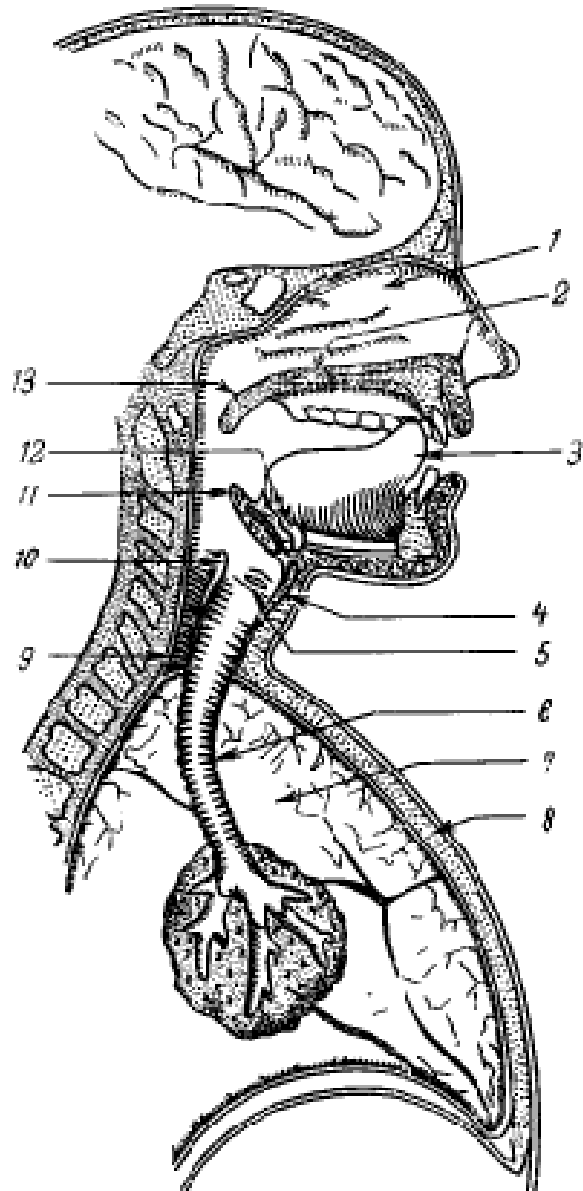
Оцінювання параметрів формант

Синтез голосних звуків

Частина 1

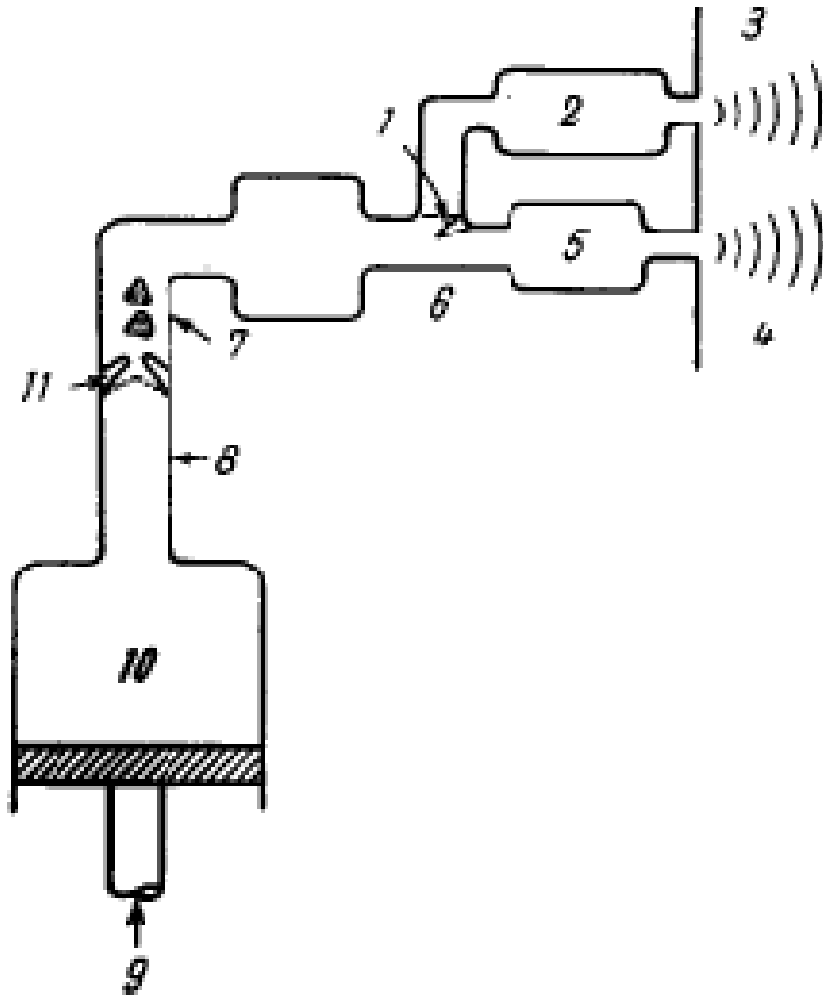
- 1. Визначення форманти**
- 2. Застосування непараметричних оцінок спектра для оцінювання параметрів формант**
- 3. Прості способи синтезу голосних звуків**

Схематичне зображення мовного апарату людини



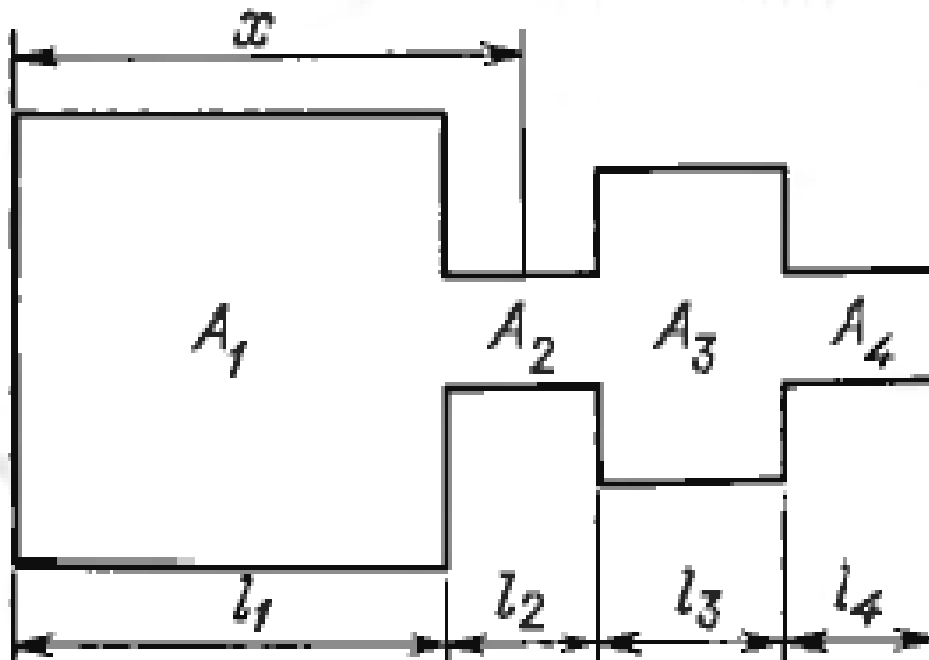
- 1 – носовая полость,
- 2 – твердое небо,
- 3 – язык,
- 4 – щитовидный хрящ,
- 5 – голосовые связки,
- 6 – трахея,
- 7 – легкое,
- 8 – грудина,
- 9 – пищевод,
- 10 – кольцеобразный хрящ,
- 11 – надгортанье,
- 12 – подъязычная кость,
- 13 – мягкое небо

Схематичне зображення функціональних частин голосового тракту



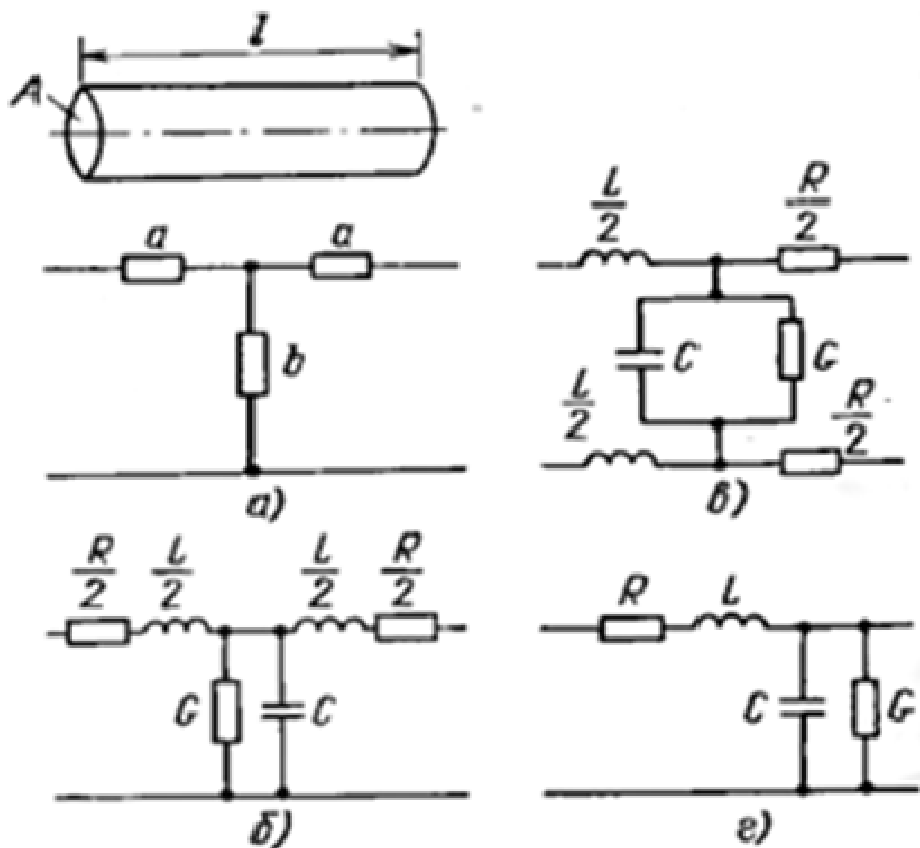
- 1 – мягкое небо,**
- 2 – носовая полость,**
- 3 – излучения носа,**
- 4 – излучения рта,**
- 5 – ротовая полость,**
- 6 – поднятая часть языка,**
- 7 – гортанная трубка,**
- 8 – трахея и бронхи,**
- 9 – мускульная сила,**
- 10 – объем легких,**
- 11 – голосовые связки**

Акустична модель системи резонаторів голосового тракту



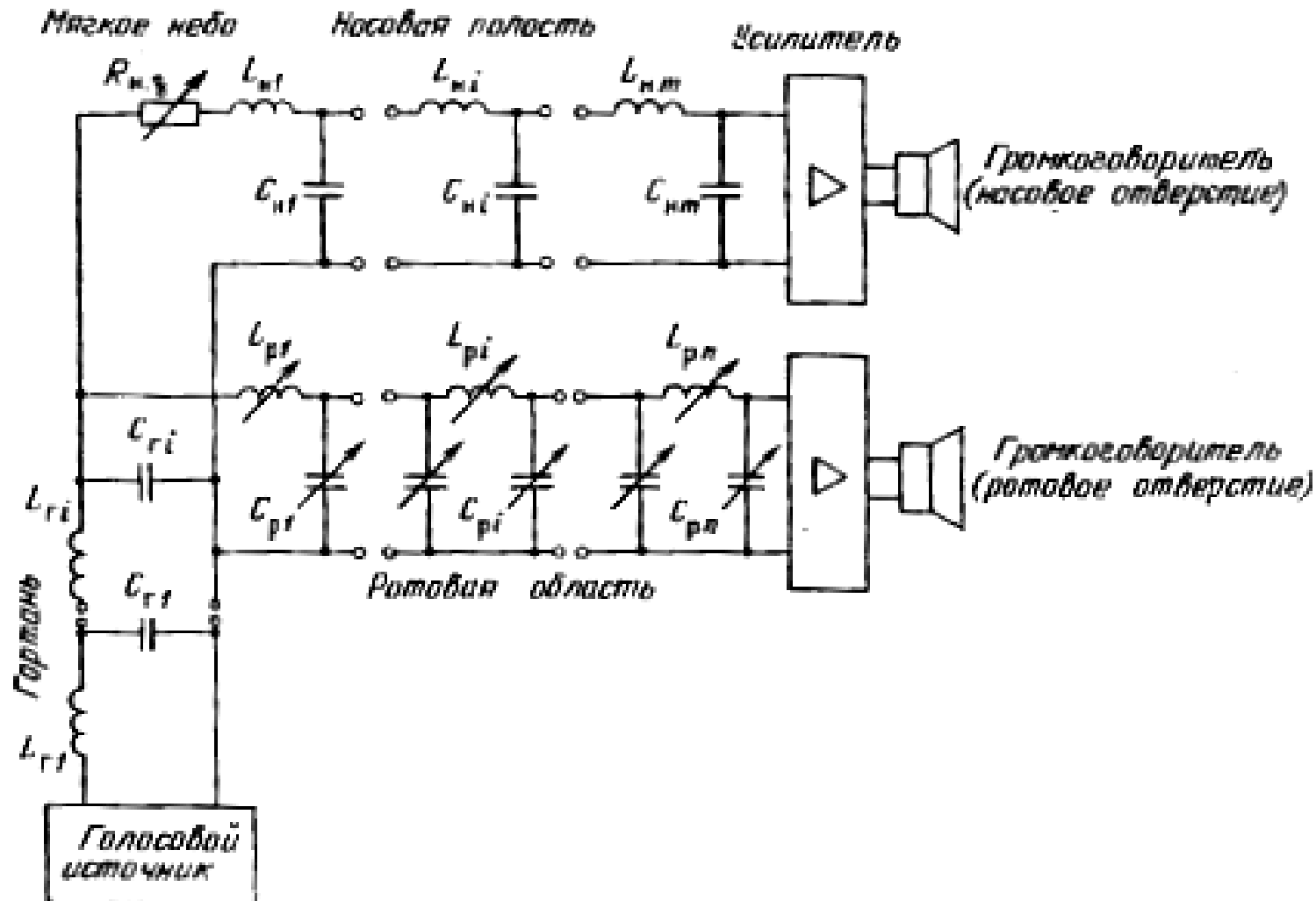
A_1 - гортань и задняя ротовая полость (до языка),
 A_2 – участок сужения между языком и твердым небом,
 A_3 – передняя ротовая полость,
 A_4 – проход между губами

Електричні еквіваленти відрізка труби



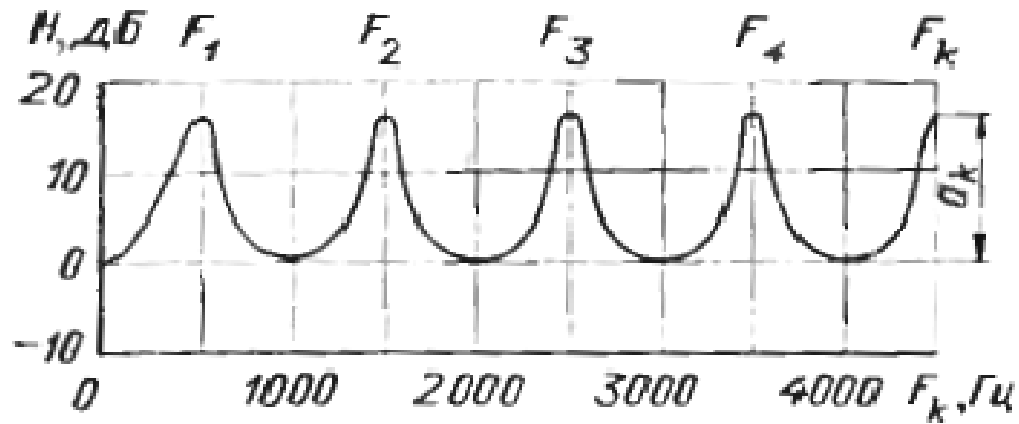
L - индуктивность
аналогична воздушной
инертности массы;
C - емкость аналогична
упругости сжимаемого
объема воздуха;
R - сопротивление
представляет потери,
обусловленные вязким
трением о стенки трубы;
G – эквивалентная
проводимость описывает
потери, вызванные
теплопроводностью
стенок трубы

Электрический эквивалент акустической модели голосового тракта



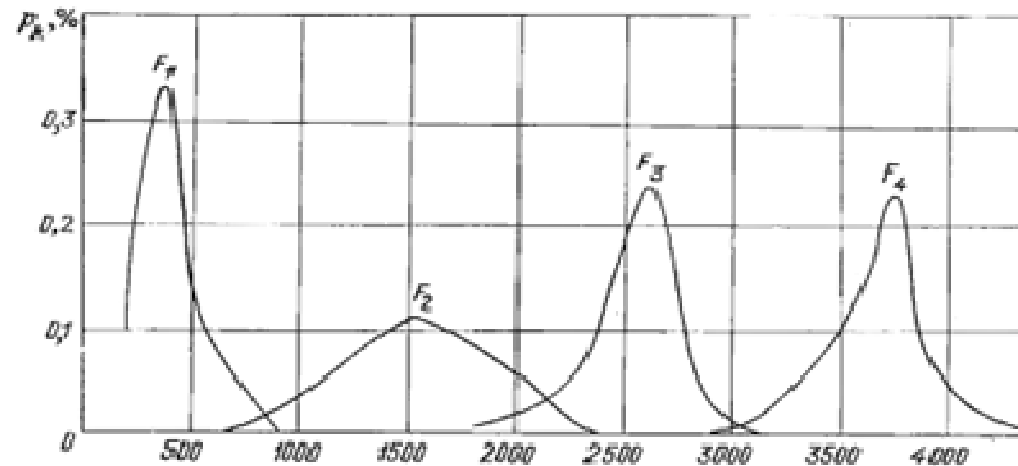
Модельні та експериментальні дані

АЧХ однострубно́ї моделі голосового апарату



Труба перетином 5 см^2 ,
довжиною $17,5 \text{ см}$

Розподіл ймовірностей формантних частот чоловічих голосів

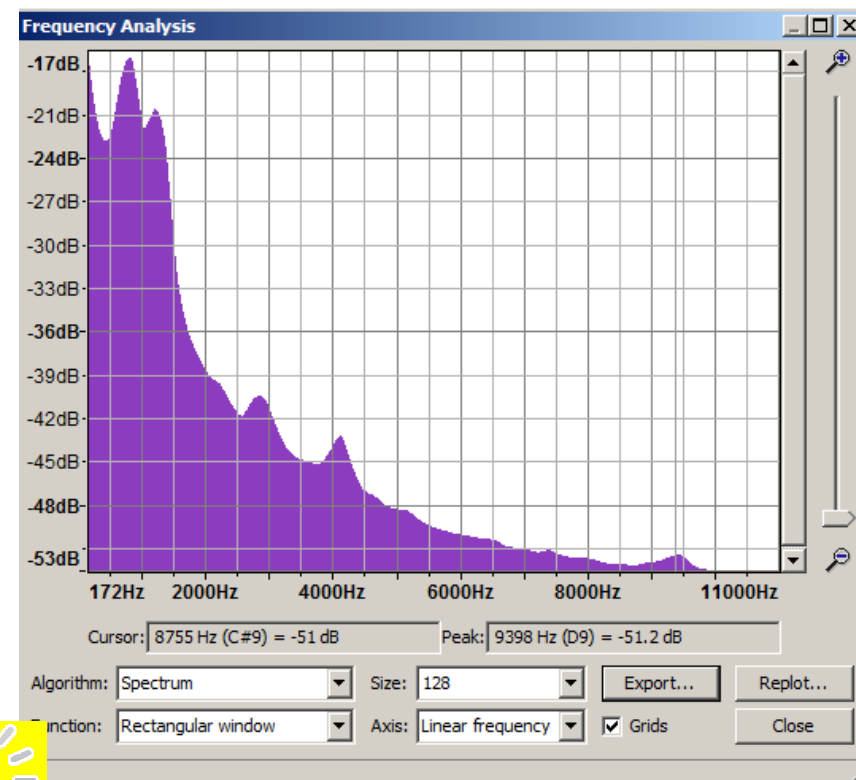
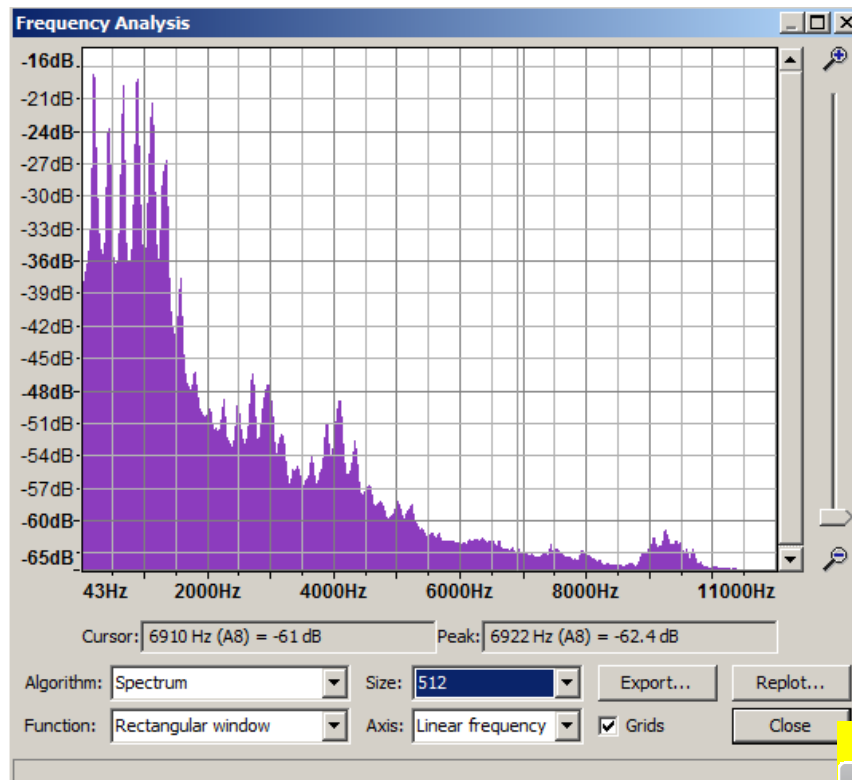


Таблиця параметрів формант російської мови

Звуки		1-я формантная полоса, Гц	2-я формантная полоса, Гц
Гласные	а	1100...1400	-
	и	2800...4200	-
	о	400...800	-
	у	200...600	-
	ы	200...600	1500...2300
	э	600...1000	1600...2500
Согласные	з	0...600	4200...8600
	ж	200...600	1350...6300
	л	200...500	700...1100
	м	0...400	1600...1850
	н	0...400	1500...3400
	р	200...1500	-
	с	4200...8600	-
	ф	7000...12000	-
	х	400...1200	-
	ш	1200...6300	-

Візуалізація формант. Спосіб 1. Непараметричні оцінки спектра із низькою роздільною здатністю за частотою

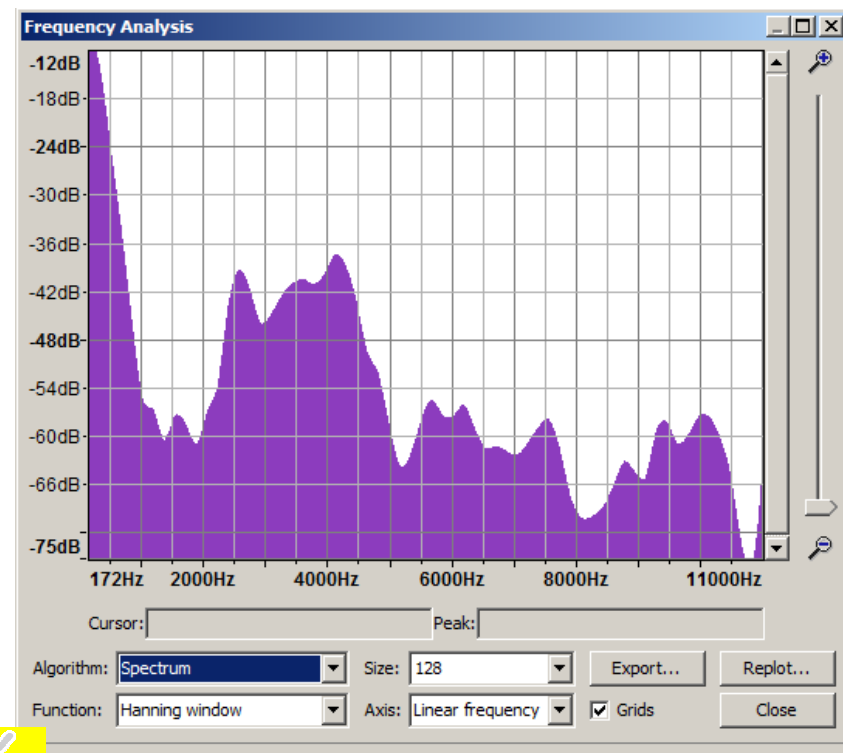
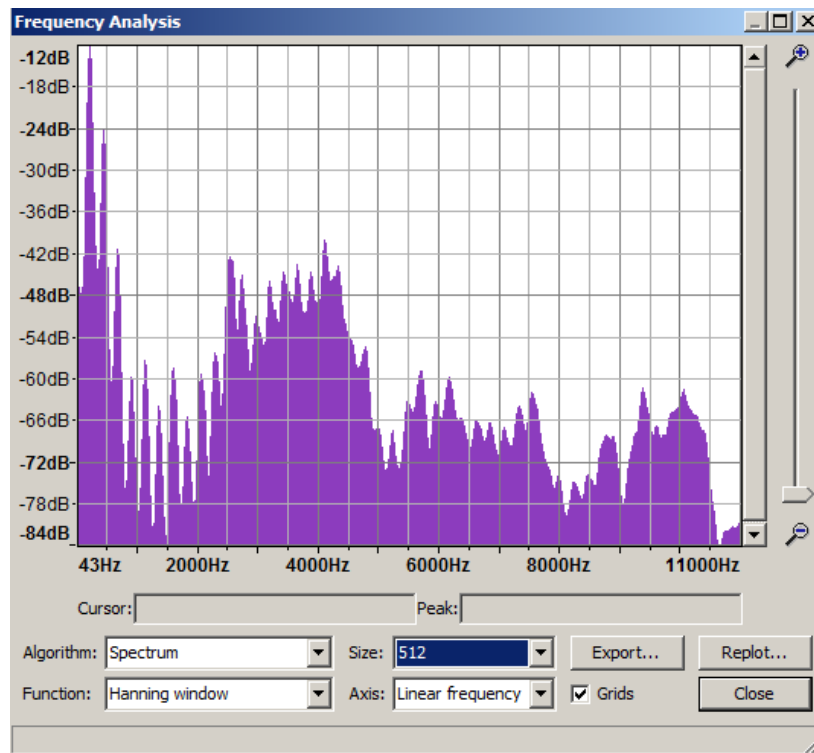
Звук «а». Жіночий голос. $N_{\text{seg}} = 512$ та $N_{\text{seg}} = 128$, $F_s = 22050$ Гц



Візуалізація формант (продовження)

Спосіб 1. Непараметричні оцінки спектра із низькою роздільною здатністю за частотою

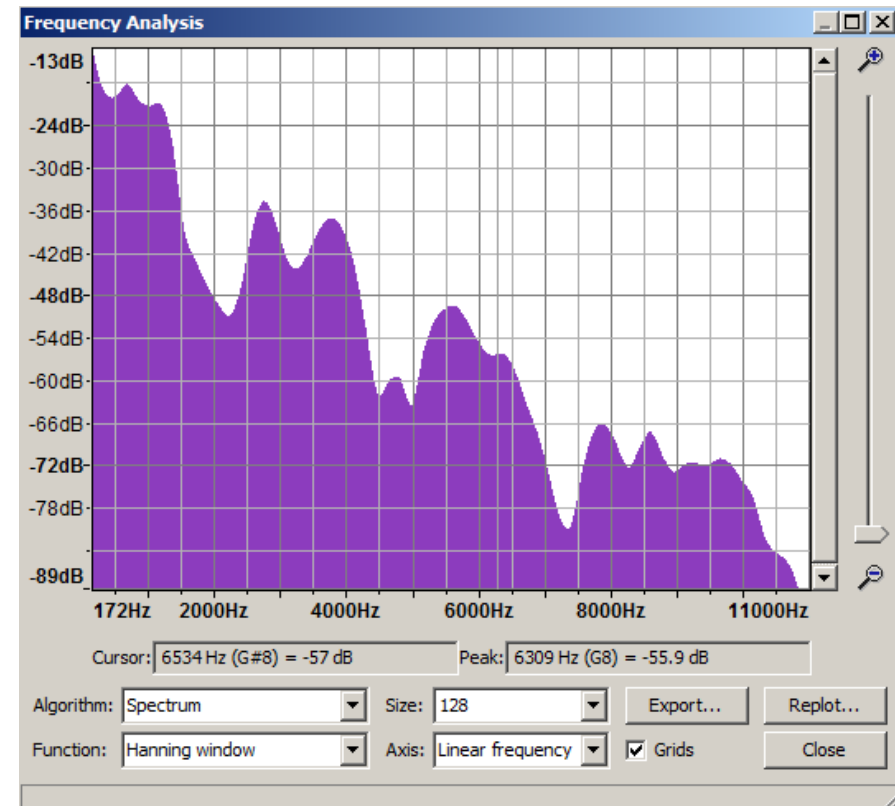
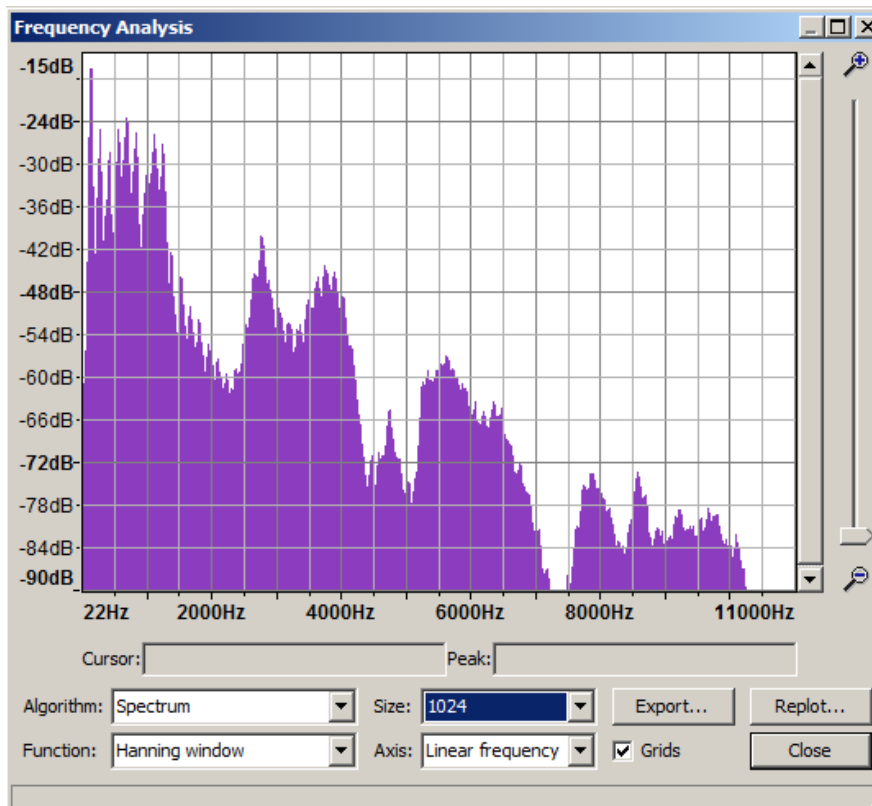
Звук «і». Жіночий голос. Nseg = 512 та Nseg = 128, Fs = 22050 Гц



Візуалізація формант (продовження)

Спосіб 1. Непараметричні оцінки спектра із низькою роздільною здатністю за частотою

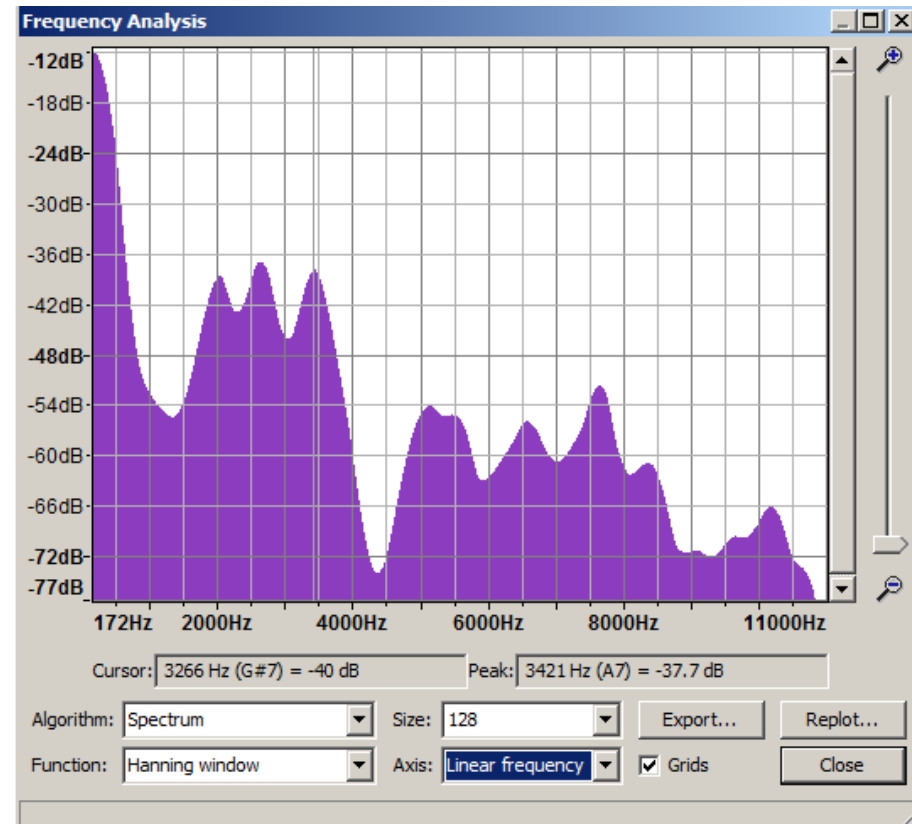
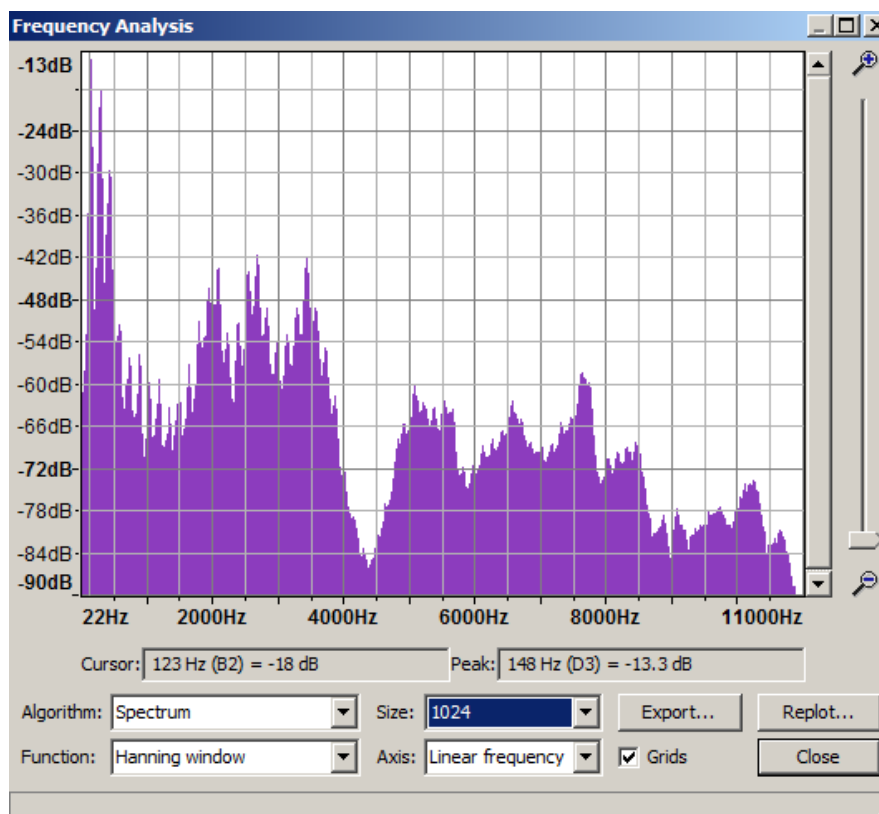
Звук «а». Чоловічий голос. $N_{\text{seg}} = 1024$ та $N_{\text{seg}} = 128$, $F_s = 22050$ Гц



Візуалізація формант (продовження)

Спосіб 1. Непараметричні оцінки спектра із низькою роздільною здатністю за частотою

Звук «і». Чоловічий голос. $N_{\text{seg}} = 1024$ та $N_{\text{seg}} = 128$, $F_s = 22050$ Гц



Оцінки параметрів формант

Результати наших вимірювань

		f0	F1	F2	F3	F4
а	Чумак (ж)	224	870	-	2870	4040
	Мазепін (ч)	135	670	1170	2770	3790
і	Чумак (ж)	232	1600	2617	3570	4140
	Мазепін (ч)	140	2040	2670	3440	7660

Літературні дані (російська)

	F1	F2
а	1100-1400	-
і	2800-4200	-

Причини розбіжності?

Аналіз причин та корегування оцінювання

- $T_{\text{seg}} = N_{\text{seg}}/F_s$; $df = 1/T_{\text{seg}} = F_s/N_{\text{seg}}$ (зазвичай $N_{\text{seg}} = N_{\text{fft}}$)
- Для $N_{\text{seg}} = 128$ та $F_s = 22050$ маємо $df = 172$ Гц
- **Висновок: роздільну здатність за частотою треба погіршити**

- Що робити, якщо Audacity не дозволяє зробити $N_{\text{fft}} < 128$?
- **Треба збільшити F_s шляхом передискретизації (інтерполяція в частотній області)**

Перевірка припущення

Результати нових вимірювань (df = 344 Гц)

		F1	F2
а	Чумак (ж)	950	2870
	Мазепін (ч)	2770	3790
і	Чумак (ж)	2617	4140
	Мазепін (ч)	2670	3440

Літературні дані (Покровський)

	F1	F2
а	1100-1400	-
і	2800-4200	-

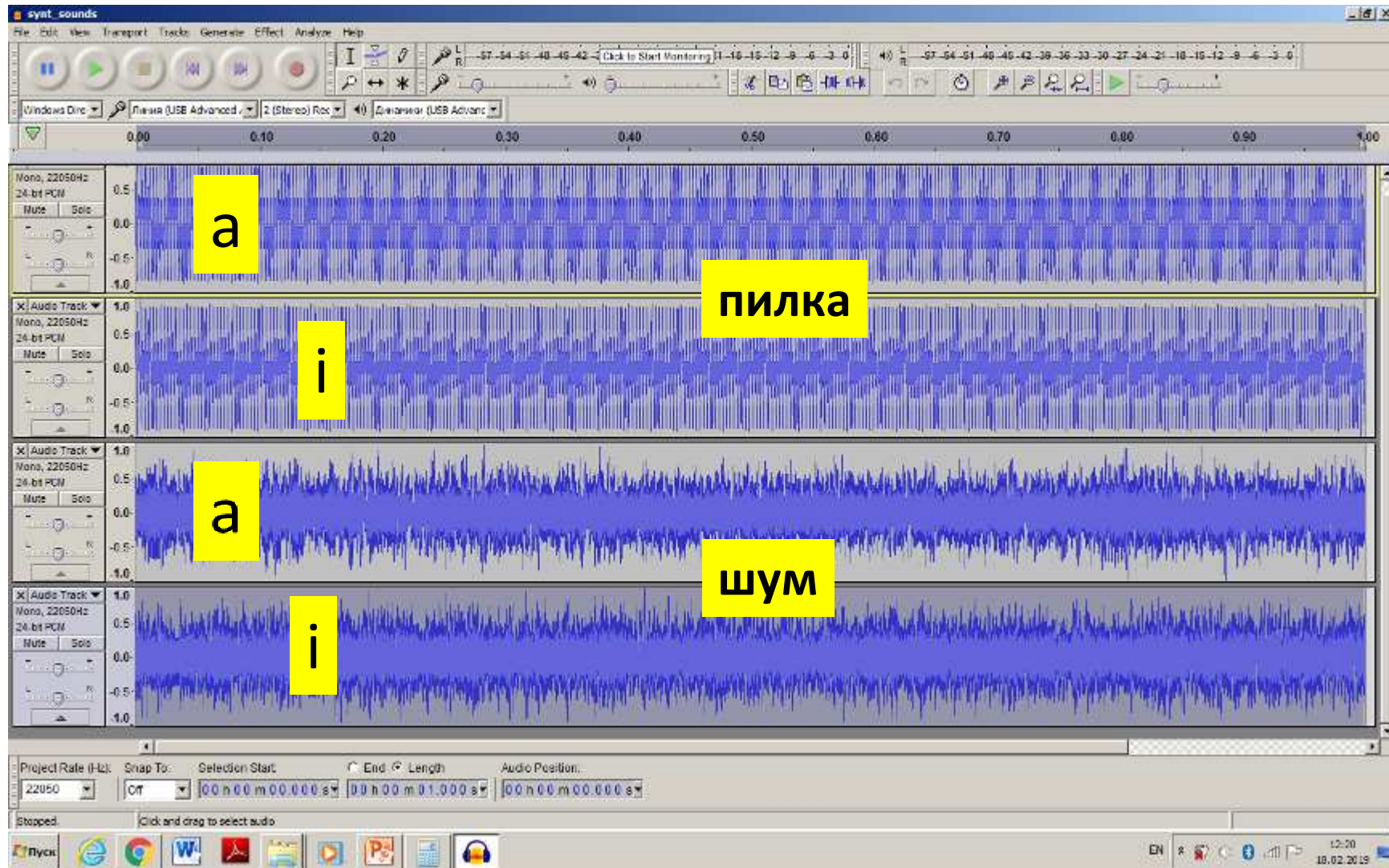
Літературні дані (англ., Wiki)

	F1	F2
а	850	1610
і	240	2400

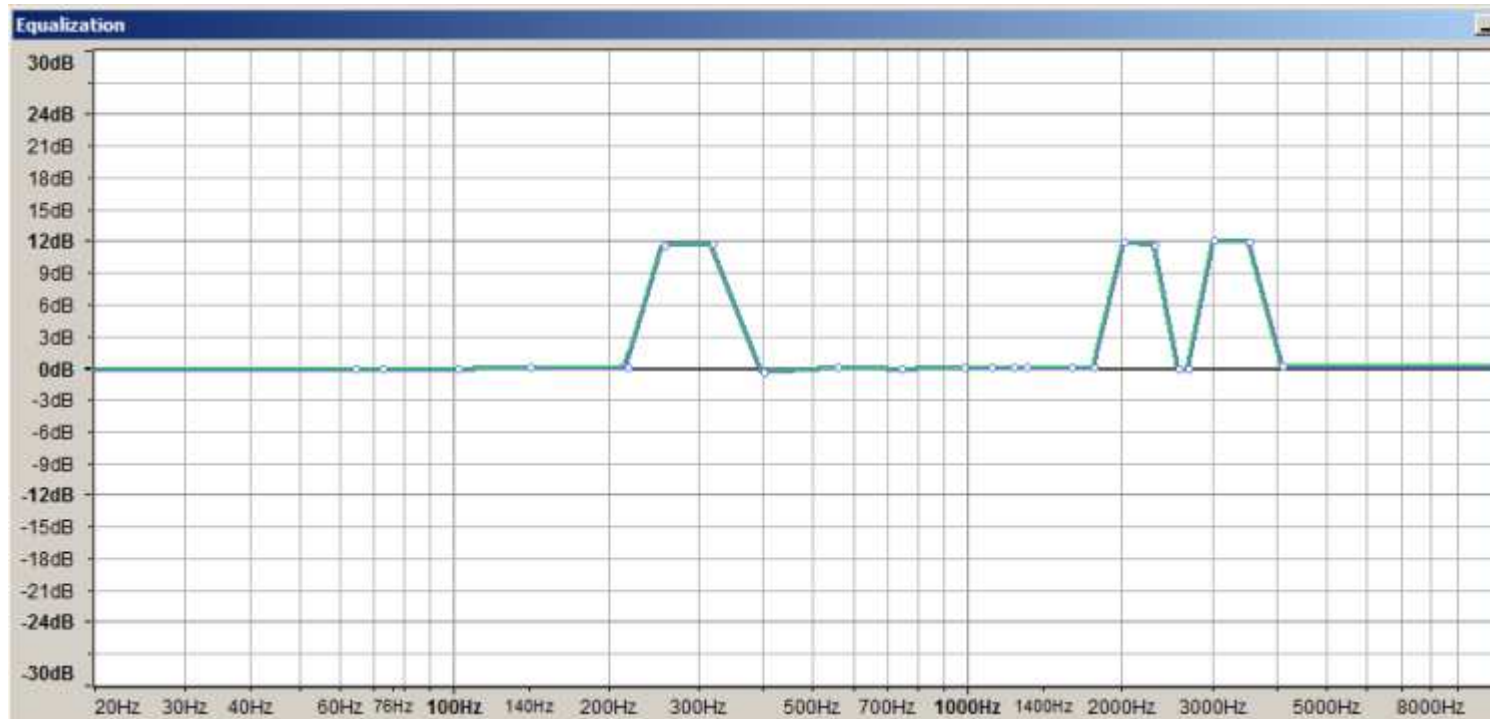
Літературні дані (англ., Duda)

%	Vowel	F1	F2	F3
%	/a/	730	1090	2440
%	/i/	270	2290	3010

Синтез голосних звуків (Audacity)



Еквалайзер як полірезонансний фільтр (Aidacity)



Синтез голосних звуків (MakeVowel)

```
Editor - D:\Acoustic\!!! COBRAS\Voices_natur_and_synt\MakeVowel.m
+3 MicBeamChirp1kHz_varBand.m x t60_shreder_2.m x testWienerSubtr.m x myspecgram.m x MakeVowel.m x +
1 function [y,y1]=MakeVowel(len, pitch, sampleRate, f1, f2, f3)
2 % MakeVowel(len, pitch [, sampleRate, f1, f2, f3]) - Make a vowel with
3 % "len" samples and the given pitch. The sample rate defaults to
4 % be 22254.545454 Hz (the native Macintosh Sampling Rate). The
5 % formant frequencies are f1, f2 & f3. Some common vowels are
6 %     Vowel   f1   f2   f3
7 %     /a/    730 1090 2440
8 %     /i/    270 2290 3010
9 %     /u/    300  870 2240
10 %
11 % The pitch variable can either be a scalar indicating the actual
12 % pitch frequency, or an array of impulse locations. Using an
13 % array of impulses allows this routine to compute vowels with
14 % varying pitch.
15 %
16 % Alternatively, f1 can be replaced with one of the following strings
17 % 'a', 'i', 'u' and the appropriate formant frequencies are
18 % automatically selected.
19 % Modified by R. Duda, 3/13/94
20 % Modified by A. Prodeus, 6/13/17
21
22 % (c) 1998 Interval Research Corporation
```



Частина 2

**Застосування параметричних оцінок спектру для
оцінювання параметрів формант**

Порівняння окремих літературних джерел

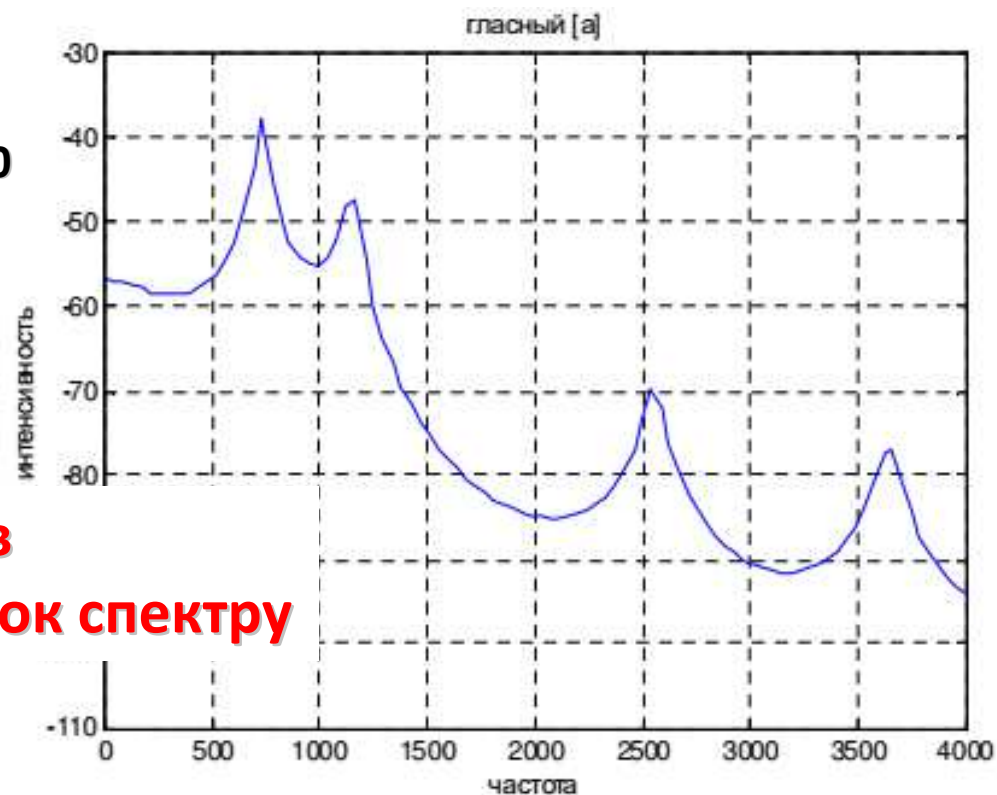
Duda R. (1994) - англ. мова

Князев С.В. (2011) – рос. мова

%	Vowel	F1	F2	F3
%	/a/	730	1090	2440

Звук /a/

1. Близькість параметрів
2. Особлива форма оцінок спектру



Видеозапись. Спектральный анализ гласного [а]

F1 = 700 Гц; F2 = 1200 Гц; F3 = 2500 Гц

Порівняння окремих літературних джерел

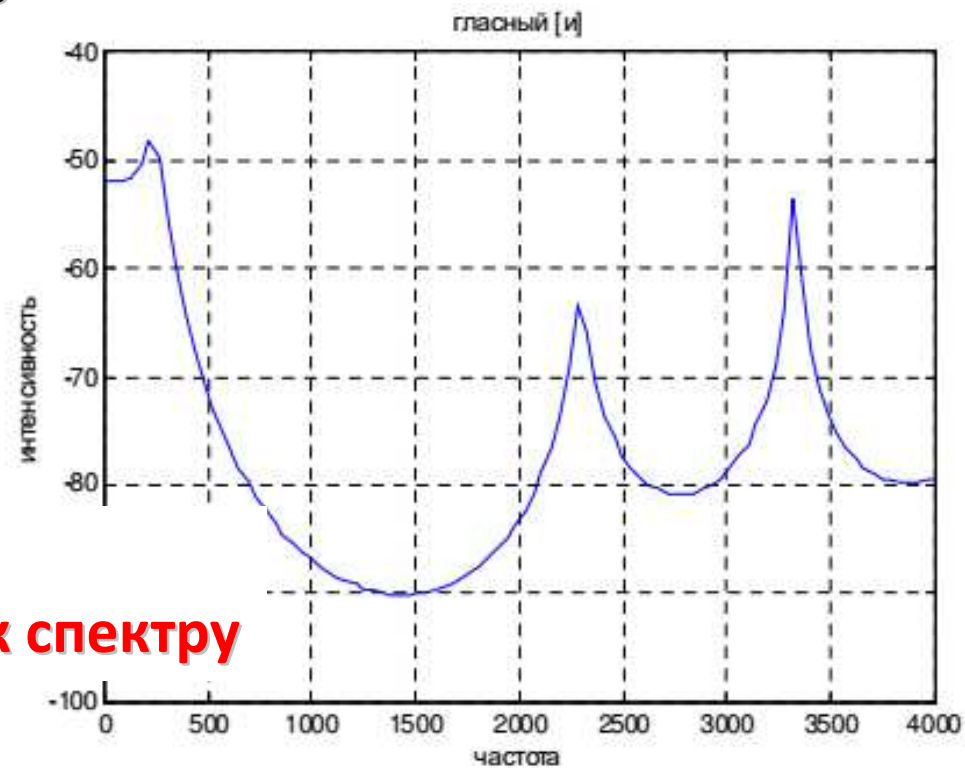
Duda R. (1994) - англ. мова

%	Vowel	F1	F2	F3
%	/i/	270	2290	3010

Князев С.В. (2011) – рос. мова

Звук /і/

1. Близькість параметрів
2. Особлива форма оцінок спектру



F1 = 200 Гц; F2 = 2250 Гц; F3 = 3300 Гц

Параметризація функцій

Ряд Тейлора:

$$f(x) = f(a) + f'(a)(x - a) + \frac{f''(a)(x - a)^2}{2!} + \dots$$

Параметри:

$$f(x) = f(a), f'(a), f''(a), \dots$$

Подання у вигляді полінома:

$$f(x) = c + bx + dx^2 + ex^3 + \dots$$

Параметри:

$$f(x) = c, b, d, e, \dots$$

Параметрична оцінка спектру

Завдання вимірювання спектру потужності еквівалентне задачі вимірювання коефіцієнтів a_k і b_k певного цифрового фільтра, на вхід якого подається білий шум, а на виході спостерігається процес із спектром потужності, що вимірюється

$$P_Y(f) = |H(f)|^2 \frac{N_0}{2}$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{A(z)}{B(z)} = \frac{\sum_{k=0}^N a_k z^{-k}}{\sum_{k=0}^M b_k z^{-k}}$$

Авторегресійний фільтр:

$$H(z) = \frac{1}{B(z)} = \frac{1}{\sum_{k=0}^M b_k z^{-k}}$$

AR оцінка Юла-Уолкера (Yule-Walker)

Система $M+1$ рівнянь:

$$\left\{ \begin{array}{l} r_n = -\sum_{k=1}^M b_k r_{n-k} \quad \text{для } 1 \leq n \leq M \\ \sigma^2 = r_0 + \sum_{k=1}^M b_k r_{-k} \quad \text{для } n = 0 \end{array} \right.$$

$r_n = \tilde{K}(n\Delta t)$ - відліки оцінки кореляційної функції процесу, що аналізується

σ^2 - оцінка дисперсії вхідного білого шуму

AR оцінка спектру Юла-Уолкера

$$\bar{P}_y(\omega) = \bar{P}_y(z) \Big|_{z=\exp(j\omega\Delta t)} = \bar{\sigma}^2 \frac{1}{\bar{B}(z)\bar{B}(z^{-1})} \Big|_{z=\exp(j\omega\Delta t)}$$

$$\bar{B}(z) = \sum_{k=0}^M \bar{b}_k z^{-k}$$

Застосування AR оцінок:

- 1) Спектр містить сплески та гладку частину
- 2) Достатньо невеликого об'єму даних

Приклади застосування AR метода:

- 1) Параметричне кодування мови (вокодери)
- 2) Розпізнавання мови

Параметричні оцінки в **sptool**

AR метод Юла-Уолкера:

Yule AR

Інші параметричні оцінки:

1) AR метод Берга (**Burg**)

2) AR коваріаційний метод (**Covariance**)

3) AR модиф. коваріац. метод (**Mod. Covar.**)

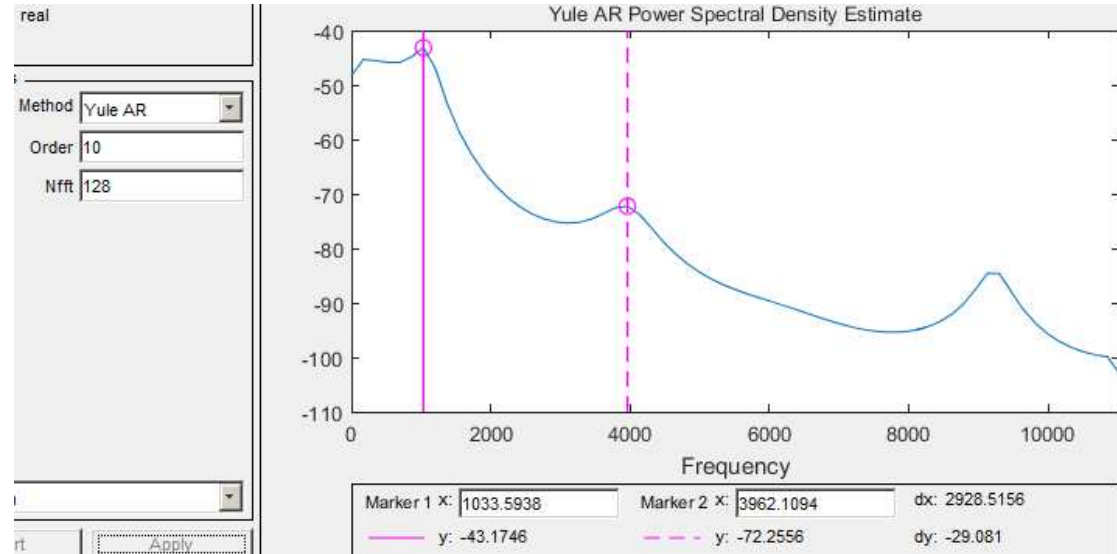
Приклад: спектр звуку /а/

Жіночий голос

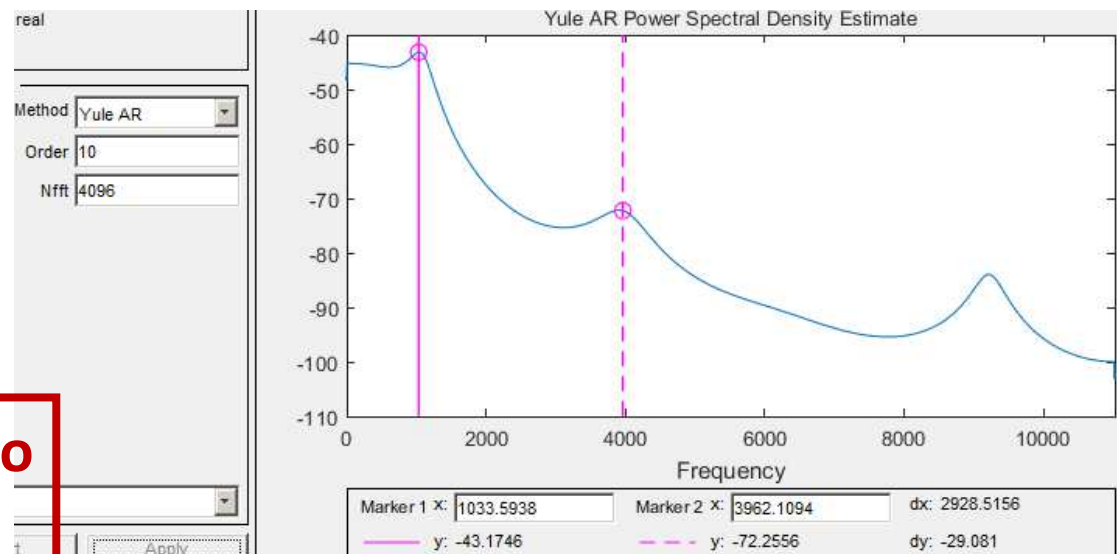
Оцінка Юла-Уолкера

Порядок $p = 10$

Nfft = 128



Nfft = 4096



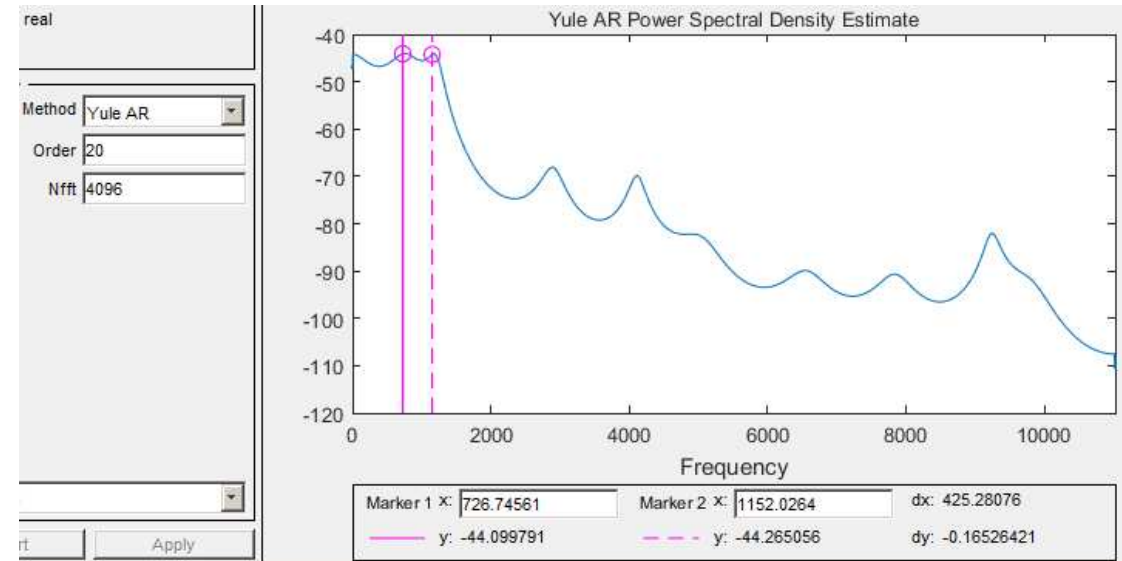
**Форма спектру практично
не залежить від Nfft**

Приклад: спектр звуку /а/ Жіночий голос

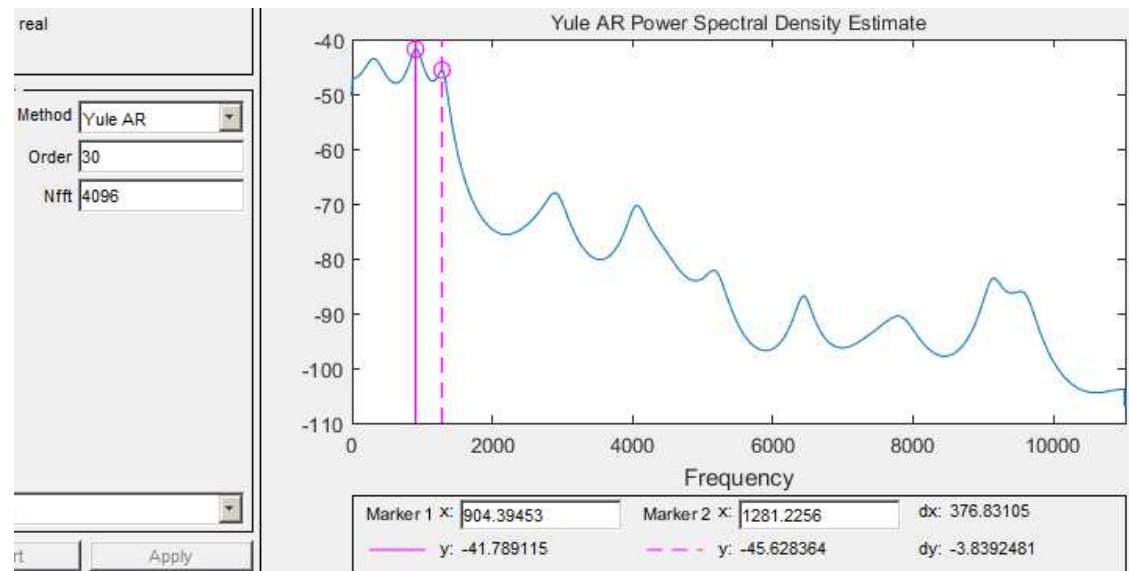
Оцінка Юла-Уолкера

Nfft = 4096

P = 20



P = 30



**Форма спектру
залежить від P**

Параметричні оцінки в Matlab

$[P_{xx},f] = \text{pyulear}(x,p,nfft,fs)$

$[P_{xx},f] = \text{pburg}(x,p,nfft,fs)$

$[P_{xx},f] = \text{pcov}(x,p,nfft,fs)$

$[P_{xx},f] = \text{pmcov}(x,p,nfft,fs)$

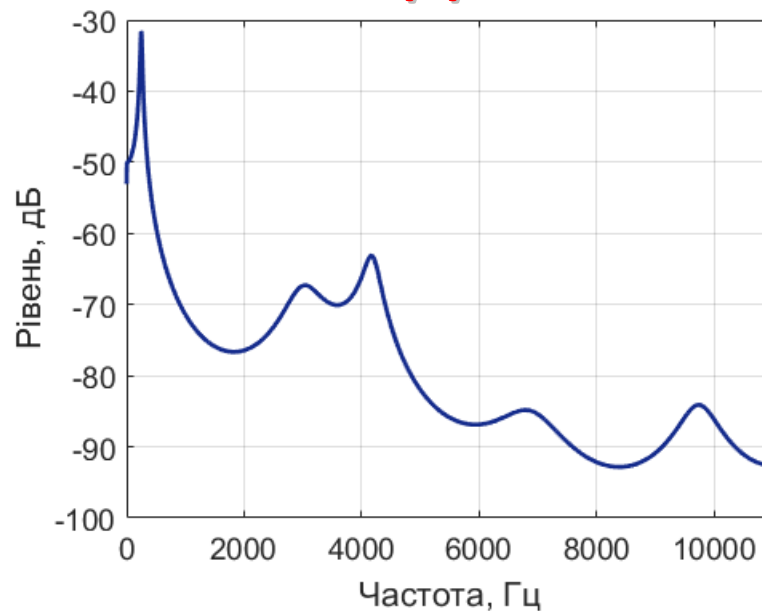
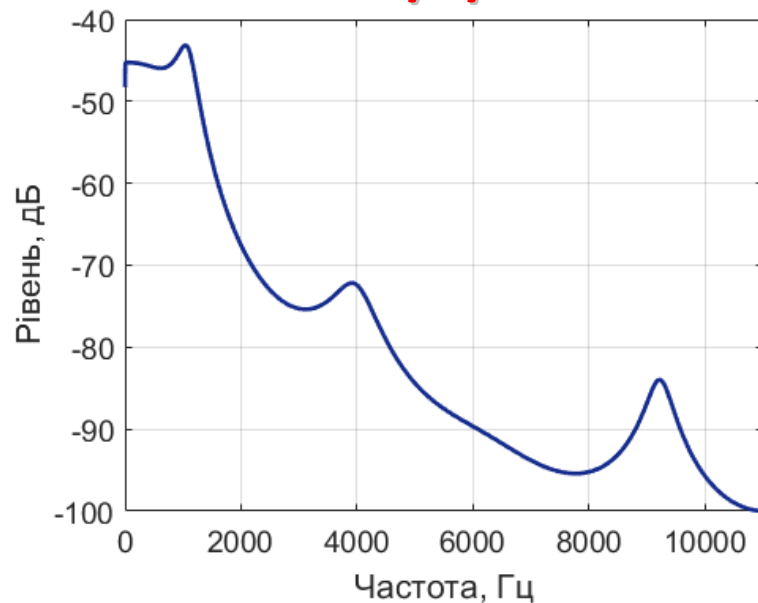
Порядок $p = 10$

p – порядок фільтра (подвоєна кількість піків спектра)

/a/

Чумак (жін.), $p = 10$

/i/

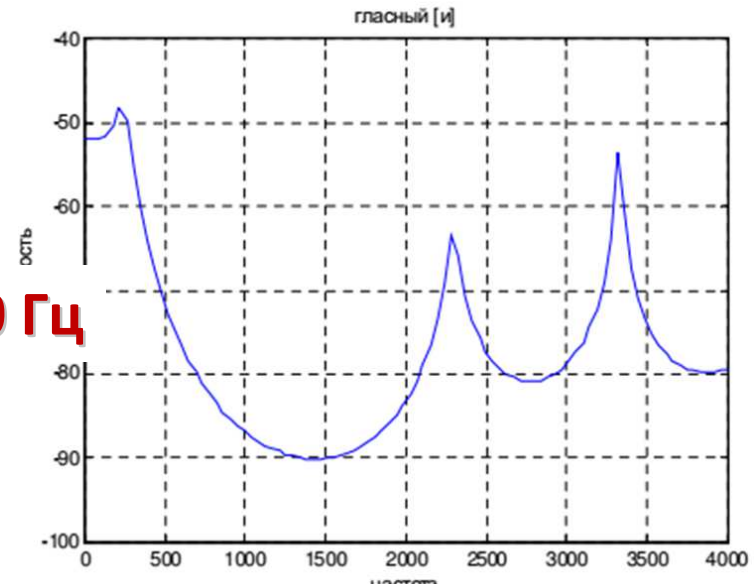


Порівняння оцінок

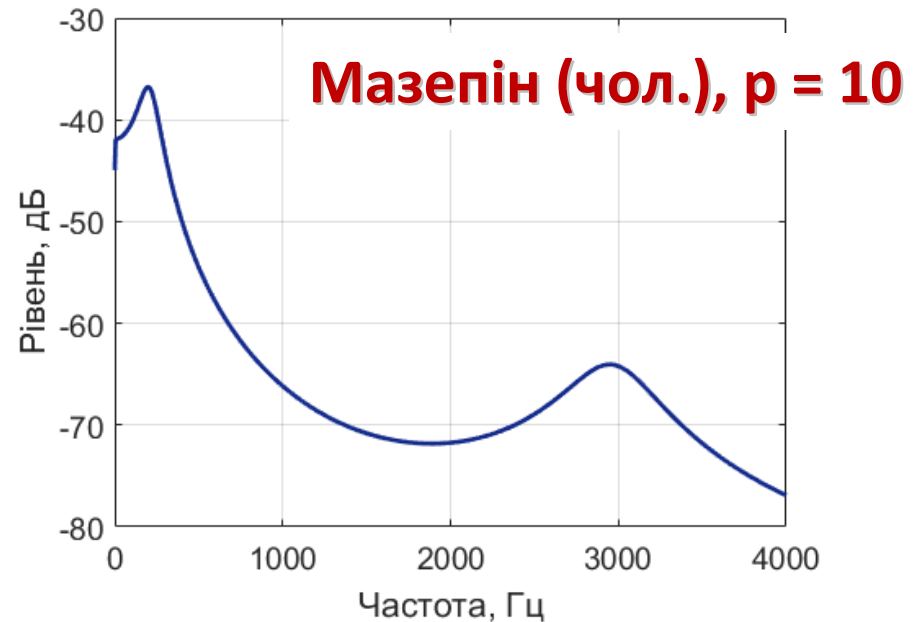
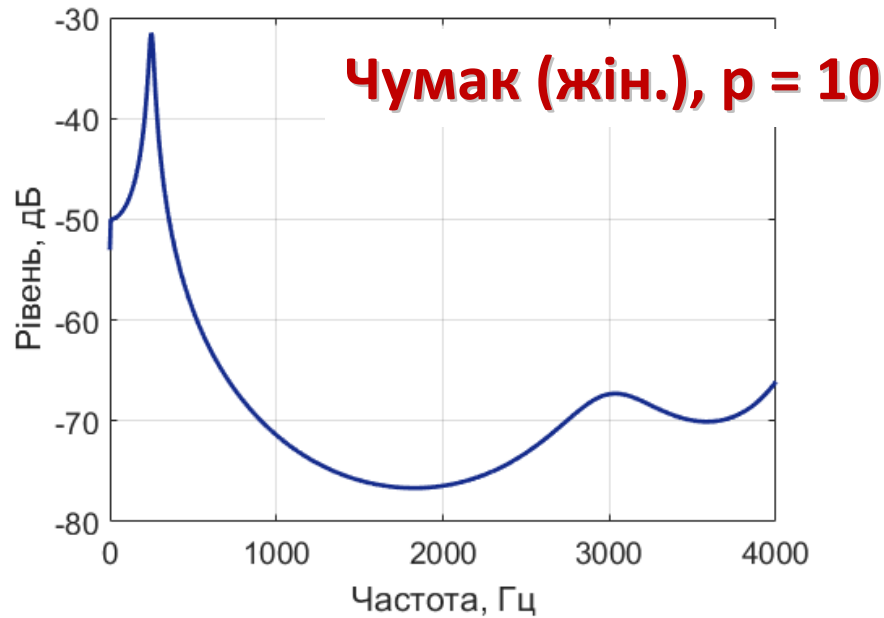
Звук /і/

Князев С.В. (2011) – рос. мова

F1 = 200 Гц; F2 = 2250 Гц; F3 = 3300 Гц



Наші оцінки



Частина 2

Застосування кепстру для формування обвідної спектру та для оцінювання параметрів формант

Оцінювання обвідної спектру через кепстр

Кепстр

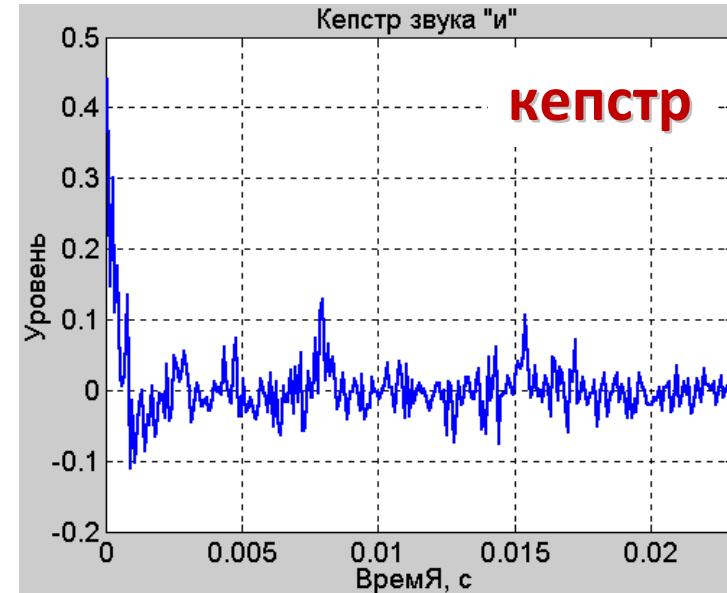
$$C(\tau, t) = F^{-1} \left\{ \ln |F_T(f, t)| \right\}$$

Обвідна спектру через кепстр

$$H(f, t) = \exp \left[F \left\{ C(\tau, t) v(\tau) \right\} \right]$$

Спектр – кепстр – обвідна спектру

$T_{фр} = 46.4$ мс; $T_{пер} = 23,2$ мс; $dF1 = 25$ Гц; $\tau = 11,2$ мс; $dF2 = 100$ Гц



Обчислення кепстра в Matlab

% === my_sepstr.m === обчислення кепстра =====

[x,Fs] = audioread('Mazepin_i.wav');

nfft = 512; fs = 11025; resample(x, fs, Fs);

tau=nfft/4+1; % ширина вікна winsepstr, що множимо на кепстр

window = rectwin(nfft); noverlap = 256;

% обчислення та графік спектрограми

[Bt,ft,tt] = spectrogram(x, window, noverlap, nfft, fs, 'yaxis');

figure; imagesc(tt,ft,20*log10(abs(Bt))), axis xy, colormap(jet);

figure; plot(ft, mean(20*log10(abs(Bt)), 2));

% ===== обчислення кепстрограми =====

Btamp = abs(Bt);

Bts=Btamp(2:end-1,:); % готуємо спектр для від'ємних частот

Btsi=flipud(Bts);

Btsim=[Btamp; Btsi]; % двосторонній спектр

Btsim=Btsim+10^(-3); % «замилуємо» нулі

Bcpstr = real(iff(log(Btsim))); % кепстрограма

tc = 0:1/fs:256/fs;

figure; imagesc(tt,tc,20*log10(abs(Bcpstr))), axis xy, colormap(jet);

figure; plot(tc, mean(Bcpstr(1:257,:), 2)); %усереднена кепстрограма

Обчислення обвідної спектру

```
% === my_spec_envel.m === Обчислення обвідної спектру =====  
% кепстральне вікно: 129 відліків + 384 нулі  
wincp = [bartlett(tau); zeros(384,1)];  
wincpstr=circshift(wincp,-(tau-1)/2); % «центрування» вікна  
wincpstr(65)=[]; % видалення зайвого відліку вікна  
  
Bcut=Bcpstr;  
for it = 1:size(Bcut, 2);  
    Bcutwin(:,it) = Bcut(:,it).*wincpstr(it);  
end  
Pcutlog = real(fft(Bcutwin));  
Pcut = exp(Pcutlog); % спектрограма обвідної спектру  
  
% графіки  
figure; imagesc(tt, ft(1:4:256), 20*log10(abs(Bcut))), axis xy, colormap(jet);  
figure; plot(ft(1:4:256), mean(20*log10(abs(Pcut(1:4:256,:))), 2));  
hold on; plot(ft, mean(20*log10(abs(Bt)), 2));  
legend('обвідна', 'спектр')
```