

Προηγμένες Τεχνικές
Επεξεργασίας Σήματος

Εργασία 4

27/5/2020

Μακρυλάκη Βασιλική-Ελένη
makrylav@ece.auth.gr
9174

Στο πρώτο κομμάτι της εργασίας το ζητούμενο είναι να ληφθούν τουλάχιστον 10 δείγματα φωνής (5 ανδρικά και 5 γυναικεία) που να περιλαμβάνουν τον ήχο των φωνηέντων (‘α’, ‘ε’, ‘ι’, ‘ο’, ‘ου’).

Όλα τα δείγματα ηχογραφήθηκαν από υπολογιστή μέσω του online-voice-recorder.com. Τα δείγματα της θηλυκής φωνής είναι από κοπέλα 22 ετών, και την αντρικής από αγόρι 27 ετών. Τα δείγματα έχουν αποθηκευτεί με ονομασίες “maleX.wav” και “femX.wav”, όπου X το αντίστοιχο φωνήεν.

Σκοπός της εργασίας είναι τα σήματα αυτά να αναπαρασταθούν στο χώρο του Cepstrum, να εκτιμηθεί η κρουστική απόκριση και μέσω αυτής να γίνει ανακατασκευή του σήματος. Για την ανακατασκευή του σήματος, οι κρουστικές αποκρίσεις συνελίχθησαν με συρμούς ώσεων $\delta[n]$.

Η υλοποίηση αυτής της διαδικασίας για κάθε σήμα, έγινε στο Matlab, μέσω της συνάρτησης `vowelProcess.m` με τον παρακάτω κώδικα:

```
function [x,period] = vowelProcess(y, fs, start, t_cut, twind,
bandwidth)

% y: the signal
% fs: data sample rate
% start: the point at which the signal is cut
% t_cut: the pitch period is included in the interval [t_cut,
t_cut+1]
% twind: the hamming window is applied in the interval [twind,
twind+3]
% bandwidth: low pass filter bandwidth / period
% x: synthesized signal
% period: pitch period in ms

m = length(y);
dt = 1/fs;
t = 0:dt:(m-1)*dt;

% Keep only 30ms of the signal
idy = (t>=start & t <= start + 0.03);
t1 = (t(idy)-start)*1000;
y1 = y(idy);
figure
plot(t1,y1);
xlabel('Time[ms]');
ylabel('Amplitude');
title('Initial signal - Time domain');
c = cceps(y1);

% Find fundamental period --- pitch
trng = t1(t1>t_cut & t1<t_cut+1);
crng = c(t1>t_cut & t1<t_cut+1);
[~,I] = max(crng);
period = trng(I);

% Apply window
trng = (t1>= twind*period & t1<= (twind+3)*period); % window length
3 periods
tnew = t1(trng);
ycut = y1(trng);
ynew = ycut.*hamming(length(tnew));

% Plot windowed result
```

```

t2 = tnew - twind*period;
figure
plot(t2,ynew)
xlabel('Time[ms]');
ylabel('Amplitude');
title('Vowel - Windowed - Time domain');

%Cepstrum
c1 = cceps(ynew);
c2 = fftshift(c1);

maxLim = max(abs(t2));
t3 = -maxLim/2:1000*dt:maxLim/2;

if length(t3) ~= length(c2')
    t3 = -maxLim/2 : 1000*dt : round((maxLim/2)*10000)/10000;
    if length(t3) ~= length(c2')
        t3 = [-round((maxLim/2)*10000)/10000: 1000*dt:
round((maxLim/2)*10000)/10000];
    end
end

figure
plot(t3 ,c2)
xlabel('Quefreny[ms]');
ylabel('Amplitude');
title('Vowel - Cepstrum domain');

% Impulse response cepstrum domain
lpf = (t3> -bandwidth*period & t3< bandwidth*period);
h = c2.*lpf';

% Impulse response Time domain
h = ifftshift(h);
H = icceps(h);
%t2 = tnew - twind*period;
figure
plot(t2,H./max(H))
xlabel('Time[ms]');
ylabel('Amplitude');
title('Impulse response - Time domain');

% ----- Synthesize back the signal -----

for i =1:length(t2)
a(i) = abs(period-t2(i));
end
[~,j] = min(a);

for i =1:length(t2)
z(i) = abs(2*period-t2(i));
end
[~,q] = min(z);

v = abs(q-j);

% Create p[n]
p(1,1:length(y)) =0;
for i =1:v:length(p)
p(i) = 1;
end

% Correlation and alignment
[w,lag] = xcorr(H,p);
[~,I] = max(w);
r = lag(I);

```

```

if r>0
p = p(r+1:end);

elseif r<0
p = p(-r+1:end);
end

% Convolution
x = conv(p,H,'same');
x = repmat(x'./max(x), round(m/length(x)),1);

t4 = 0:dt:(length(x)-1)*dt;

% Plot 30 ms of the synthesized signal
idx = (t4>=0.2 & t4<=0.23);
t5 = (t4(idx)-0.2)*1000;
s = x(idx);

figure
plot(t5,s)
xlabel('Time [ms]');
ylabel('Amplitude');
title('Synthesized signal');

```

Για την επεξεργασία κάθε φωνήεντος υλοποιήθηκαν 10 αρχεία, με όνομα 'maleXXX.m' και 'femXXX.m', όπου X το φωνήεν, στο καθένα από τα οποία καλείται η `vowelProcess.m`. Στο τέλος κάθε τέτοιου αρχείου δημιουργείται ένα νέο αρχείο ήχου με το κατασκευασμένο φωνήεν, μέσω της `audiowrite`.

Ενδεικτικά, για το θηλυκό 'α':

femAAA.m

```

clear all
close all
clc

[y,fs] = audioread('femA.wav');
[x,period] = vowelProcess(y,fs, 0.1, 4, 0.5, 0.83);
audiowrite('femA_constructed.wav', x, fs);

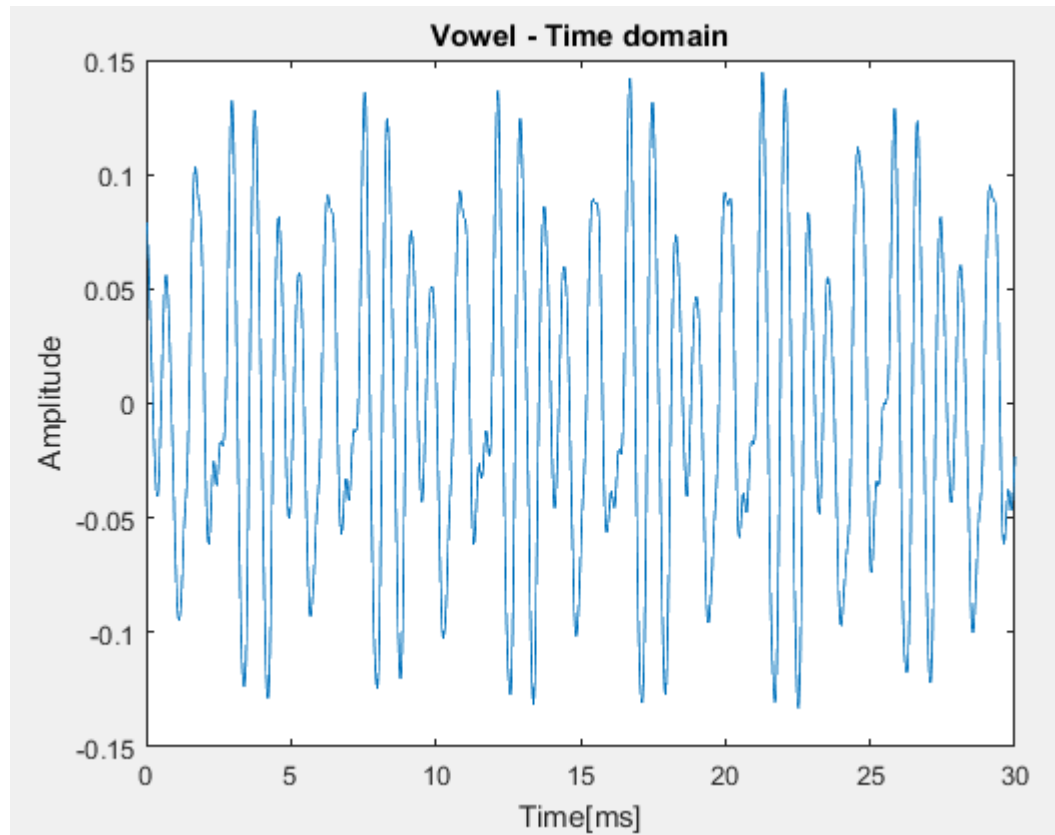
```

Σε όλες τις επεξεργασίες χρησιμοποιήθηκε Hamming window μήκους 3 periods, τοποθετημένο κατάλληλα.

Φωνήεν 'α'

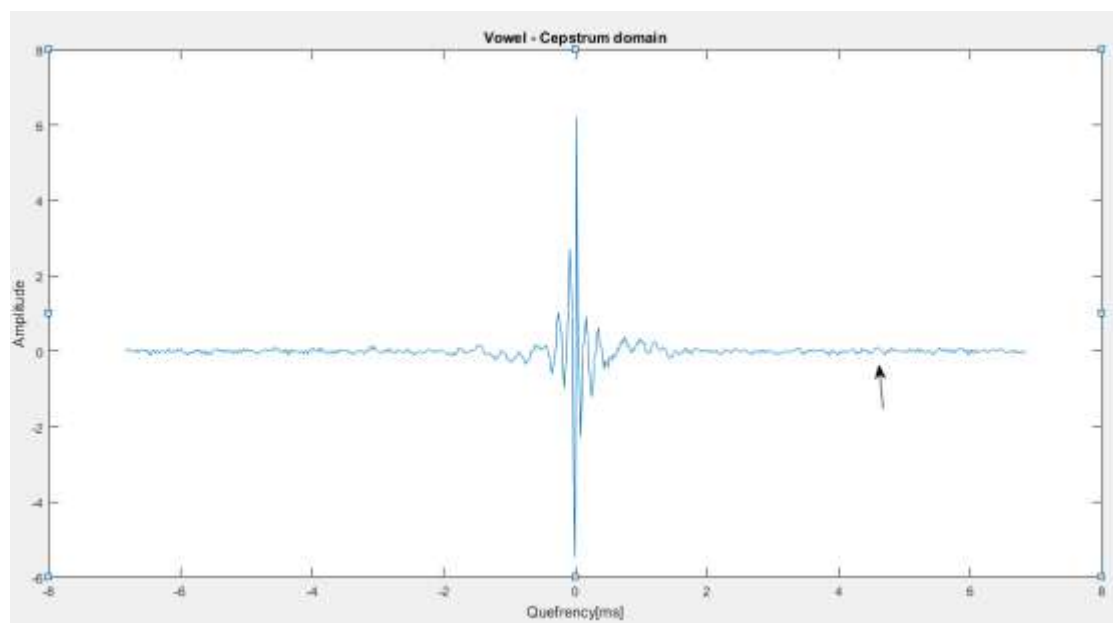
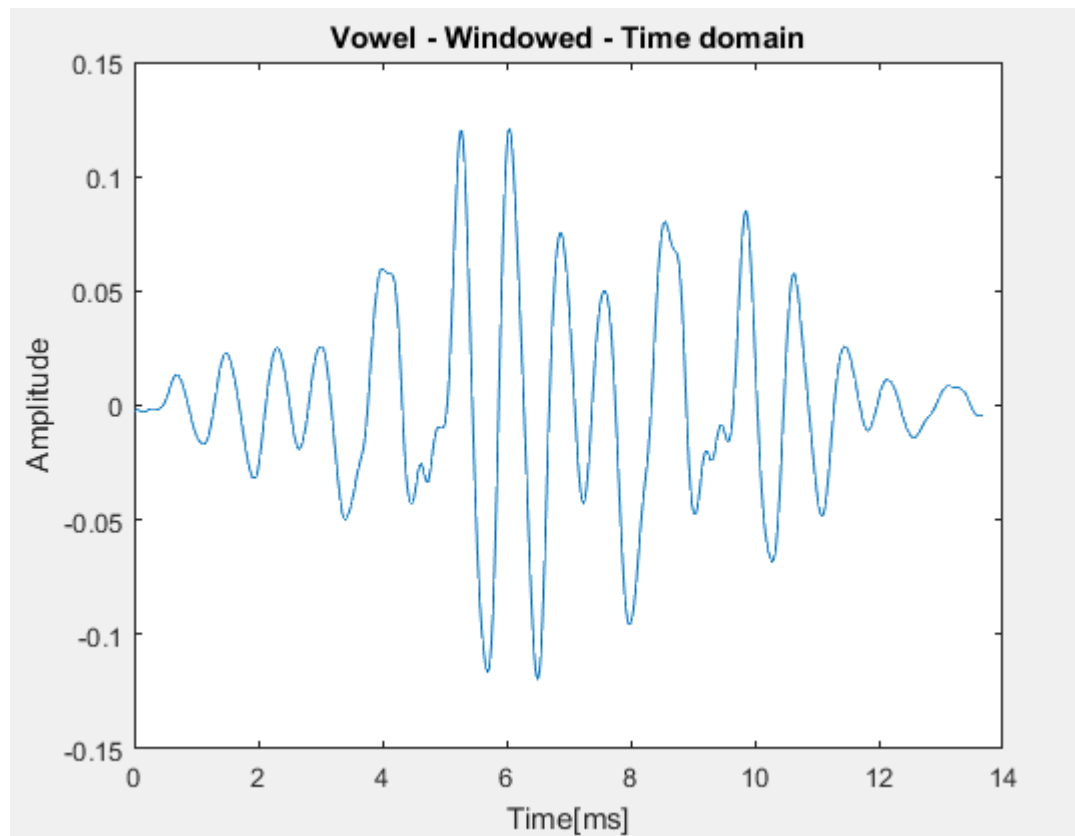
Γυναικείο 'α'

Τρέχοντας το femAAA.m, το πρώτο γράφημα περιέχει το σήμα, αναπαριστώντας μόνο 30ms αυτού, ώστε να γίνει ξεκάθαρη η μορφή του σήματος και της περιοδικότητάς του:



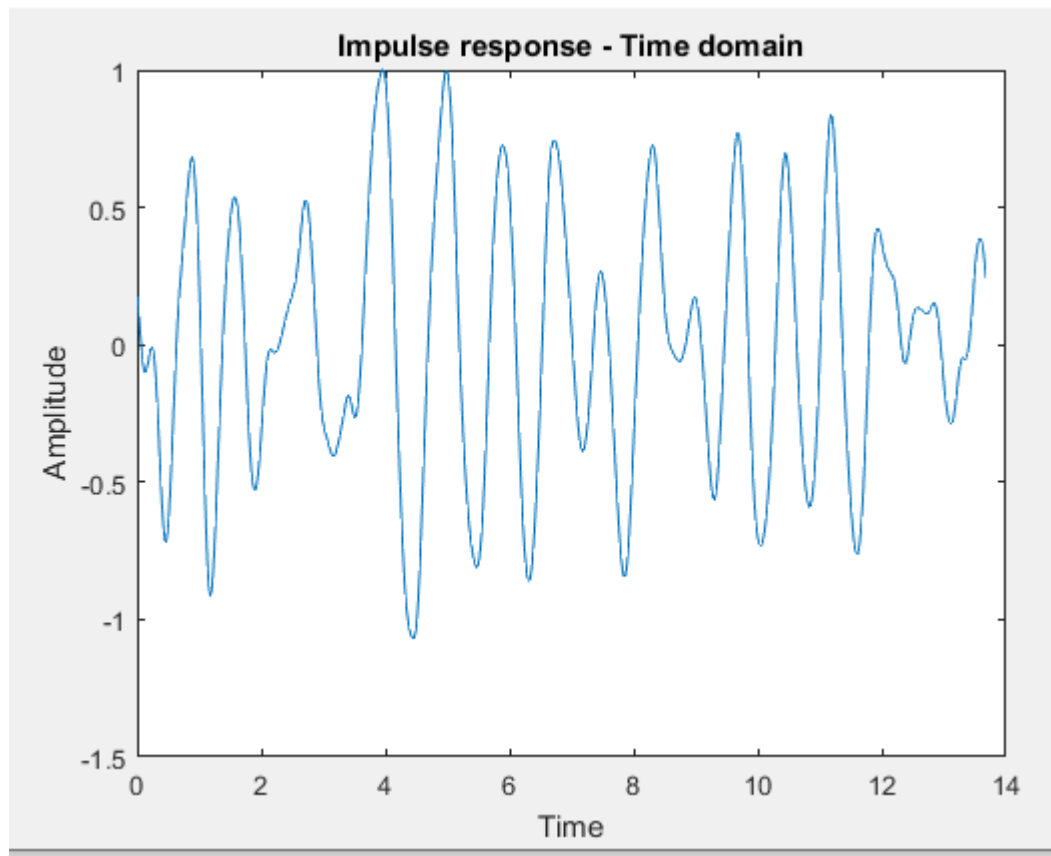
Από αυτήν την εικόνα μπορεί να εκτιμηθεί ότι η θεμελιώδης περίοδος είναι περίπου 4.56ms (~219Hz). Με βάση αυτό γίνεται και η ακριβής εκτίμησή της στον τομέα του Cepstrum καταλήγοντας ότι είναι 4.5625ms (219.178Hz).

Στη συνέχεια παρουσιάζονται τα διαγράμματα του σήματος στο χρόνο μετά την εφαρμογή του Hamming window, και στο πεδίο του Cepstrum αντίστοιχα.



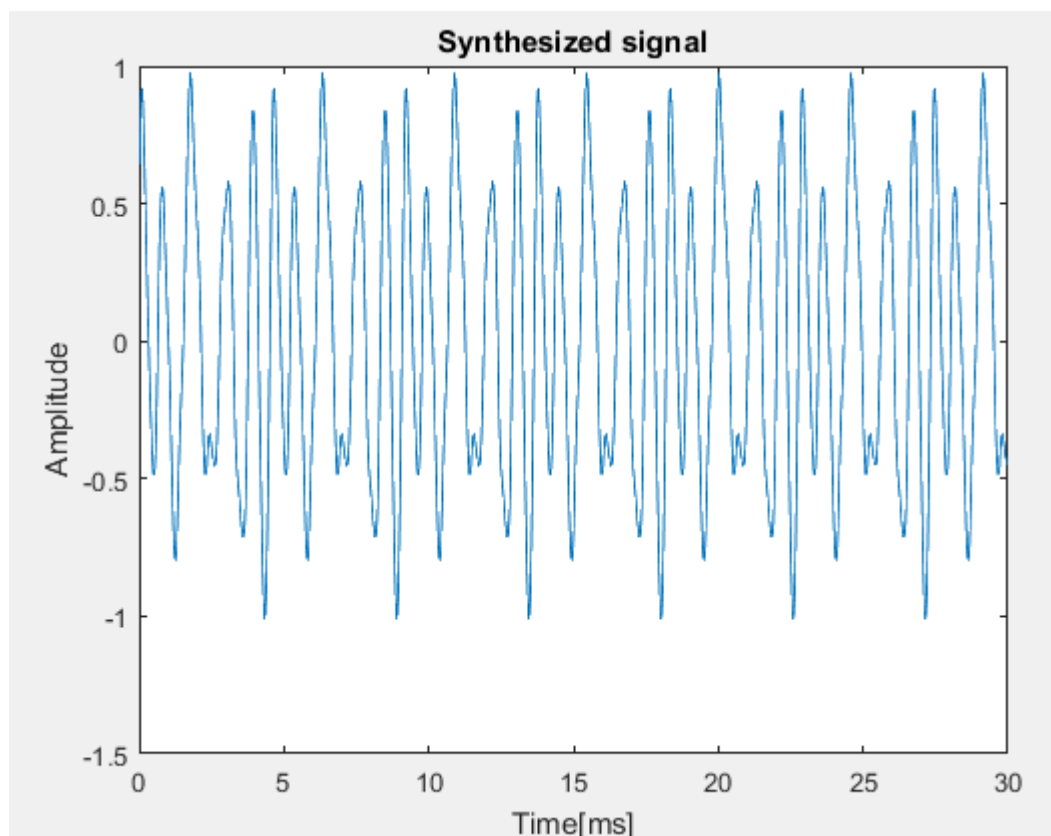
Το σημείο όπου δείχνει το βέλος αντιστοιχεί σε quefrency ίση με τη θεμελιώδη περίοδο και η κορυφή είναι πιο έντονη, όπως αναμένεται.

Έπειτα εφαρμόζεται στο σήμα χαμηλοπερατό φίλτρο μήκους 0.83 periods και γίνεται αποσυνέλιξη με σκοπό την εξαγωγή της κρουστικής απόκρισης, η οποία αναπαρίσταται στο πεδίο του χρόνου ως εξής:



Το σήμα που προκύπτει θυμίζει αρκετά αυτό που προέκυψε μετά την παραθύρωση, αλλά μετατοπισμένο.

Συνελίσσοντας το σήμα αυτό με έναν συρμό ώσεων $\delta[n]$, αφού πρώτα ευθυγραμμιστούν, προκύπτει μια ανακατασκευή του αρχικού φωνήεντος που παρατίθεται παρακάτω σε διάρκεια 30ms.

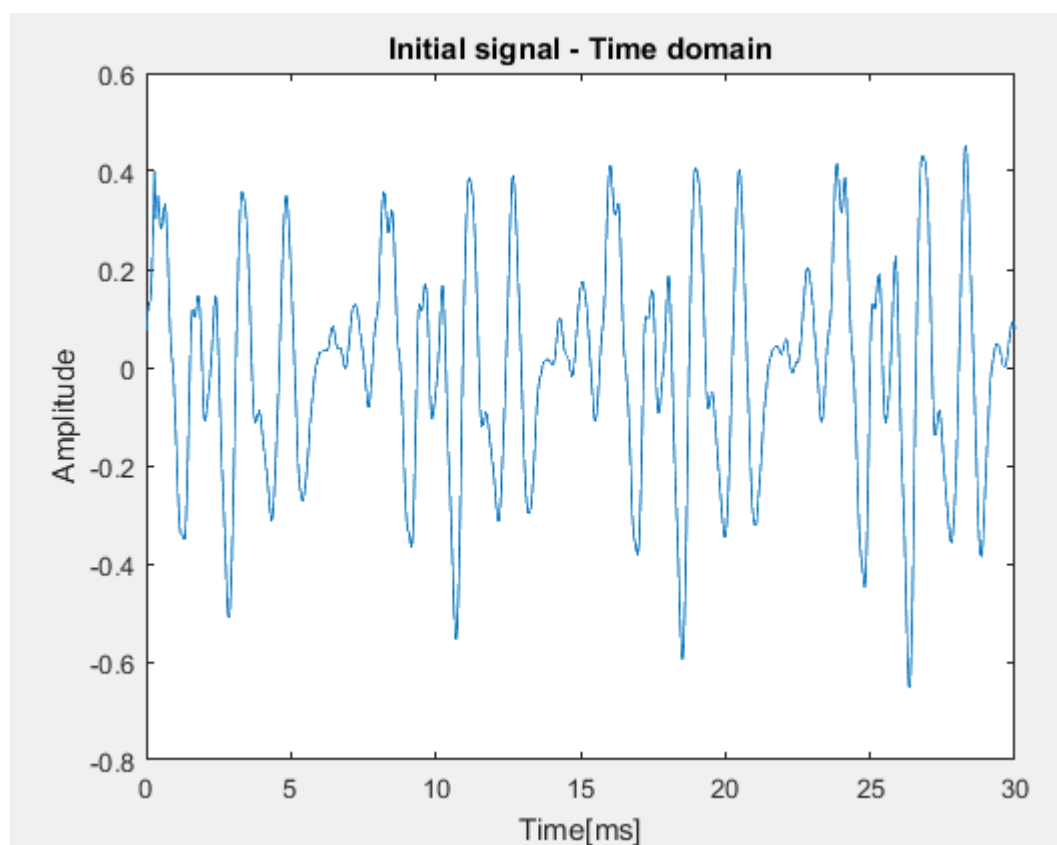


Το σήμα που προκύπτει, εκτιμάται ότι έχει μια περιοδικότητα περίπου 4.564ms, δηλαδή σχεδόν ίδια με το αρχικό σήμα, ενώ παράλληλα οι μορφές τους είναι αρκετά κοντά.

Ακούγοντας τα αρχεία ήχου femA.wav και femA_constructed.wav (ή κάνοντας χρήση στο Matlab των εντολών `sound(y, fs)`, `sound(x, fs)`), παρατηρείται ότι το κατασκευασμένο σήμα είναι σχετικά καλή προσέγγιση του πρωτότυπου, χωρίς ωστόσο να θυμίζει ανθρώπινη φωνή.

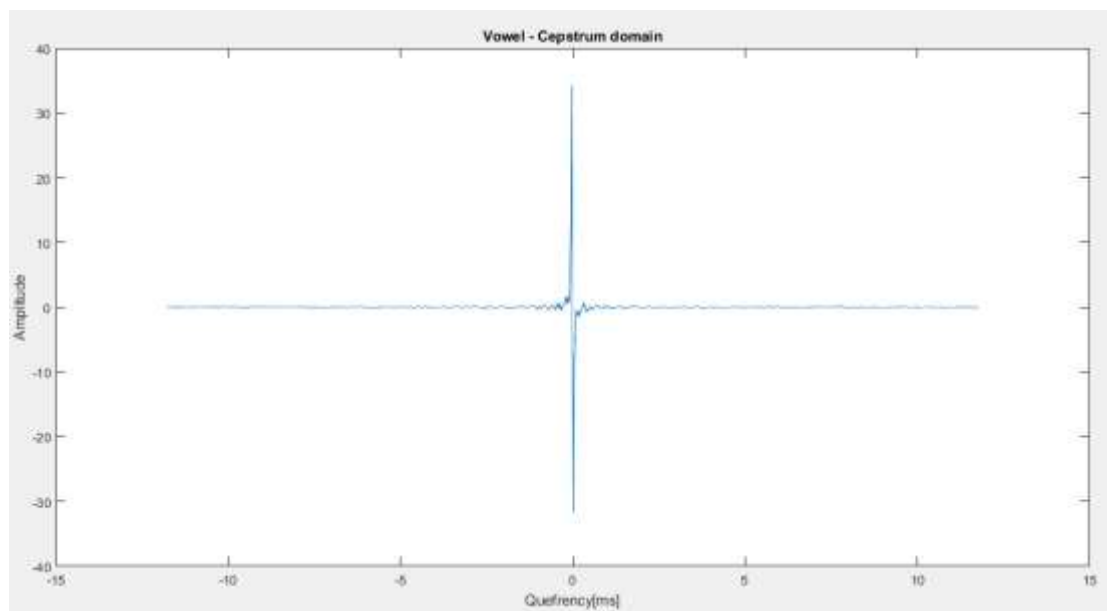
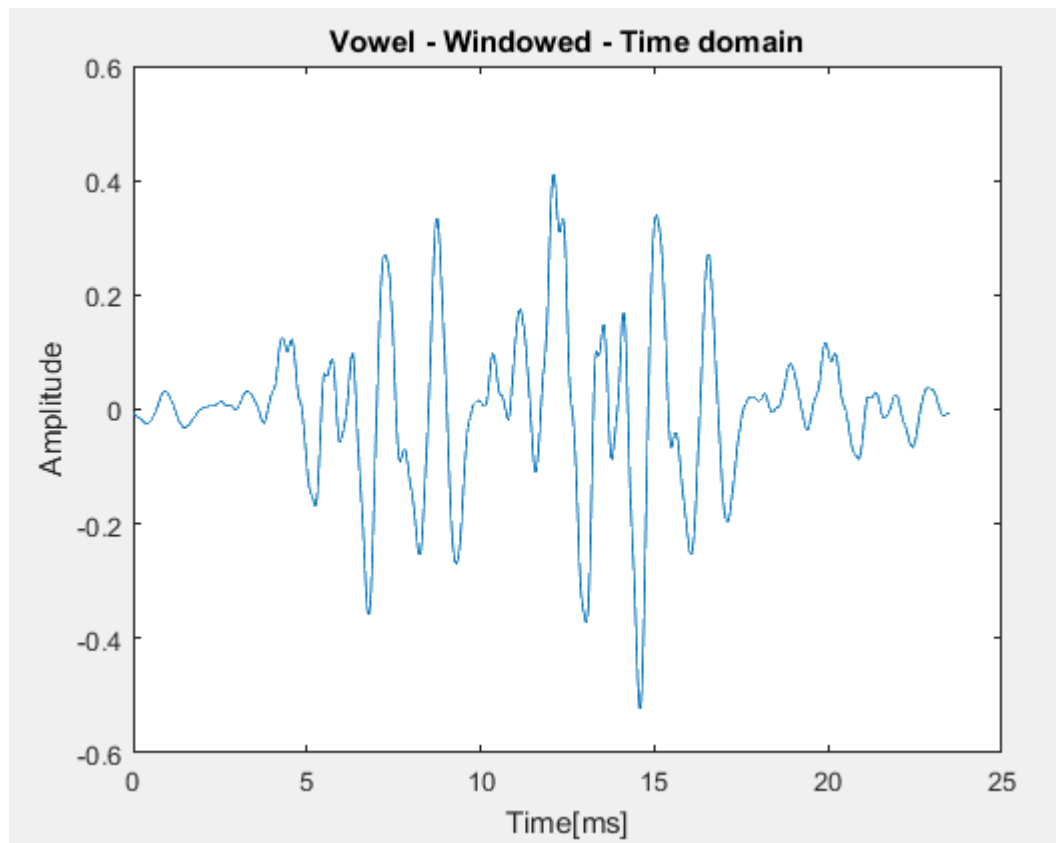
Ανδρικό 'α'

Για το ανδρικό φωνήεν 'α' τρέχουμε το αρχείο maleAAA.m και προκύπτει το εξής σήμα:



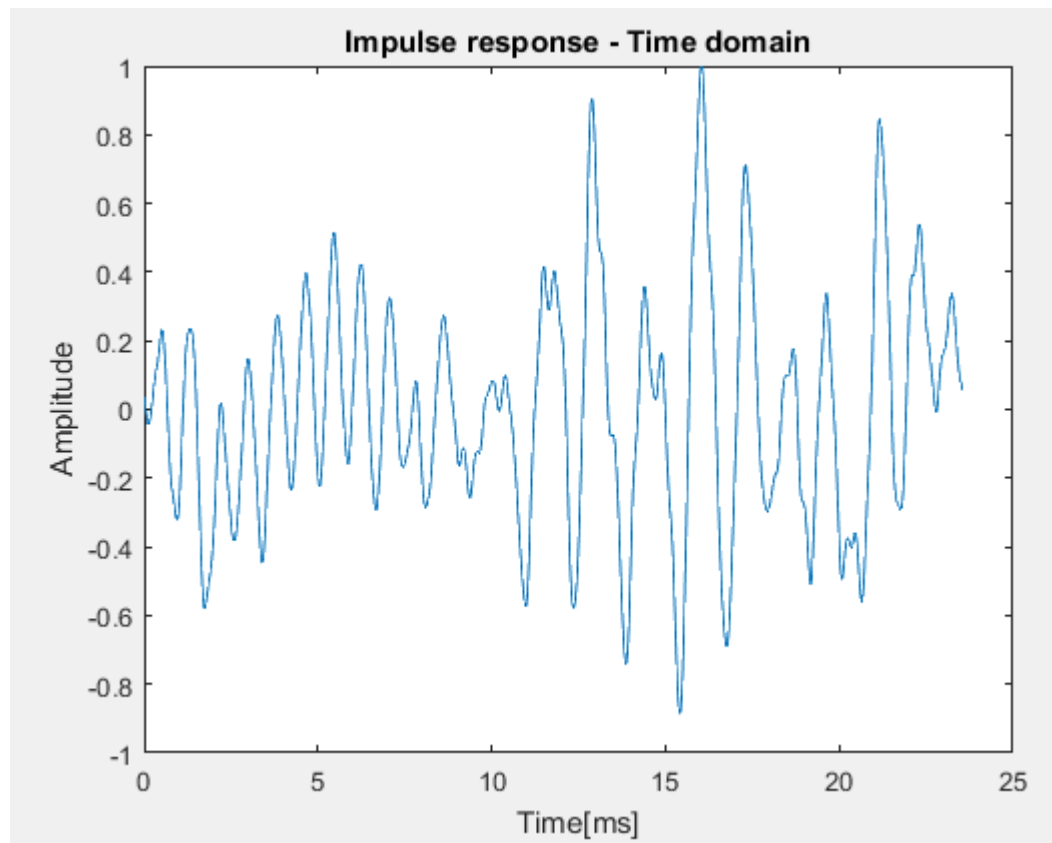
Από την εικόνα αυτή, η θεμελιώδης περίοδος εκτιμάται περίπου 7.81ms (128.041Hz) και από τον τομέα του Cepstrum προκύπτει 7.8542ms (127.32Hz, δηλαδή περίπου 92 Hz λιγότερα σε σχέση με τη θεμελιώδη συχνότητα της γυναίκας για το συγκεκριμένο φωνήεν).

Μετά την εφαρμογή του παραθύρου και την αναπαράσταση του σήματος στον τομέα του Cepstrum προκύπτουν οι εξής κυματομορφές:



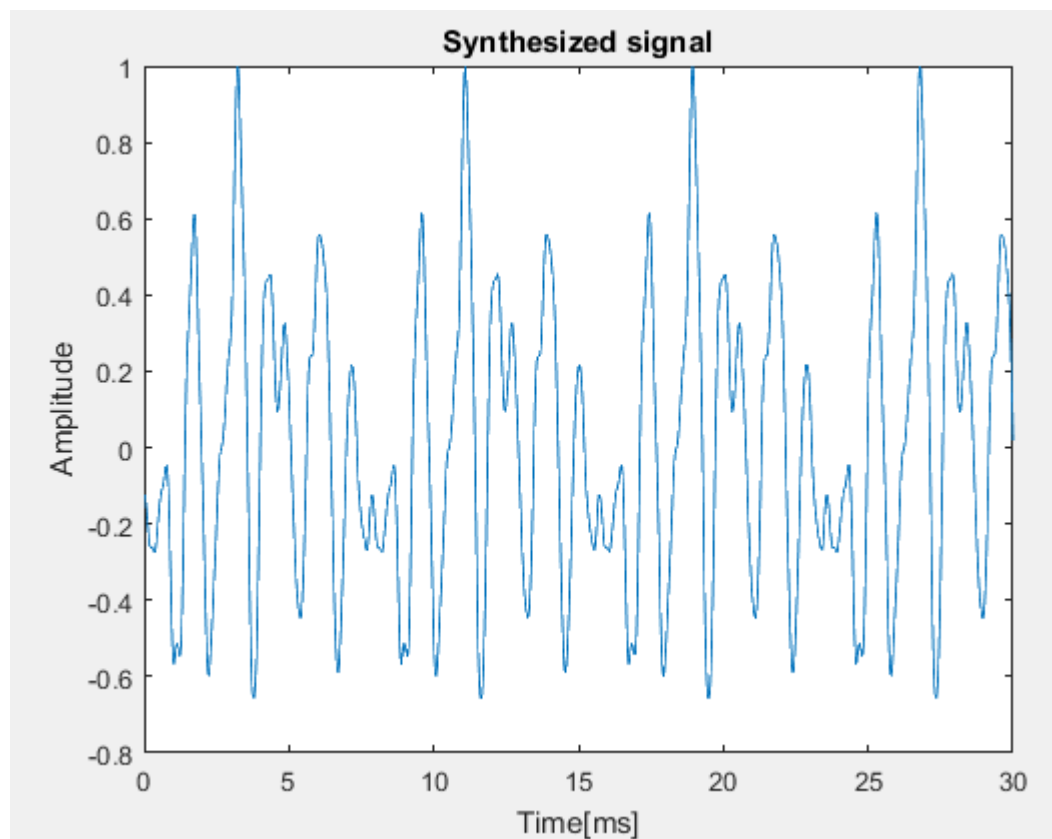
Είναι φανερό ότι στον τομέα του Cepstrum δεν υπάρχουν τόσες κορυφές όσες στο αντίστοιχο του γυναικείου 'α'. Ωστόσο, κάνοντας zoom in στα 7.85ms διακρίνεται μια κορυφή που δηλώνει το pitch.

Εφαρμόζοντας χαμηλοπερατό φίλτρο μήκους 0.62 periods και αποσυνέλιξη προκύπτει η κρουστική απόκριση που αναπαρίσταται στο χρόνο ως εξής:



Και εδώ η κρουστική απόκριση θυμίζει το σήμα που προέκυψε μετά την εφαρμογή του παραθύρου, αλλά μετατοπισμένο.

Τέλος, μετά τη συνέλιξη με το συρμό ώσεων $\delta[n]$ προκύπτει το νέο σήμα – φωνήεν:



Το σήμα αυτό εκτιμάται ότι έχει περιοδικότητα ίση με 7.86ms (δηλαδή όση και το πρωτότυπο) και παρά τη διαφορά στο πλάτος με το αρχικό σήμα, φαίνεται να είναι μια σχετικά καλή προσέγγισή του.

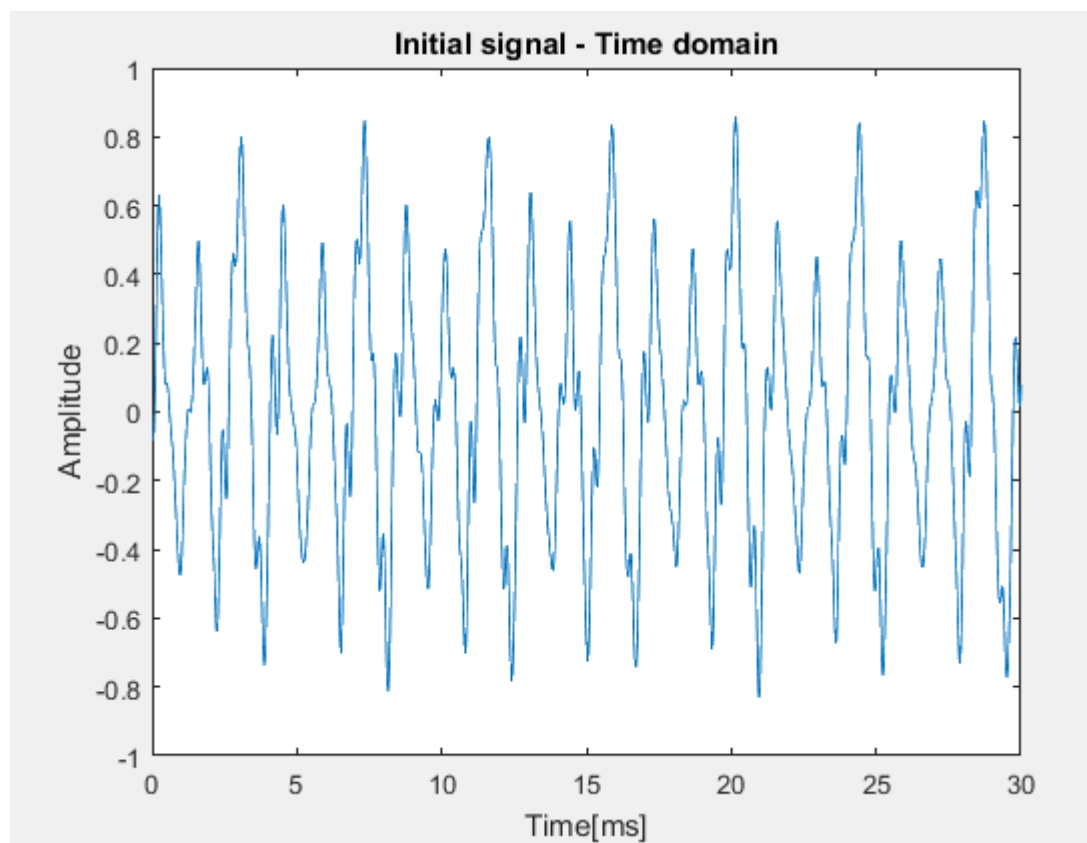
Ακούγοντας τα αρχεία maleA.wav και maleA_constructed.wav, το δεύτερο θυμίζει αρκετά την αρχική φωνή.

Φωνήεν 'ε'

Γυναικείο 'ε'

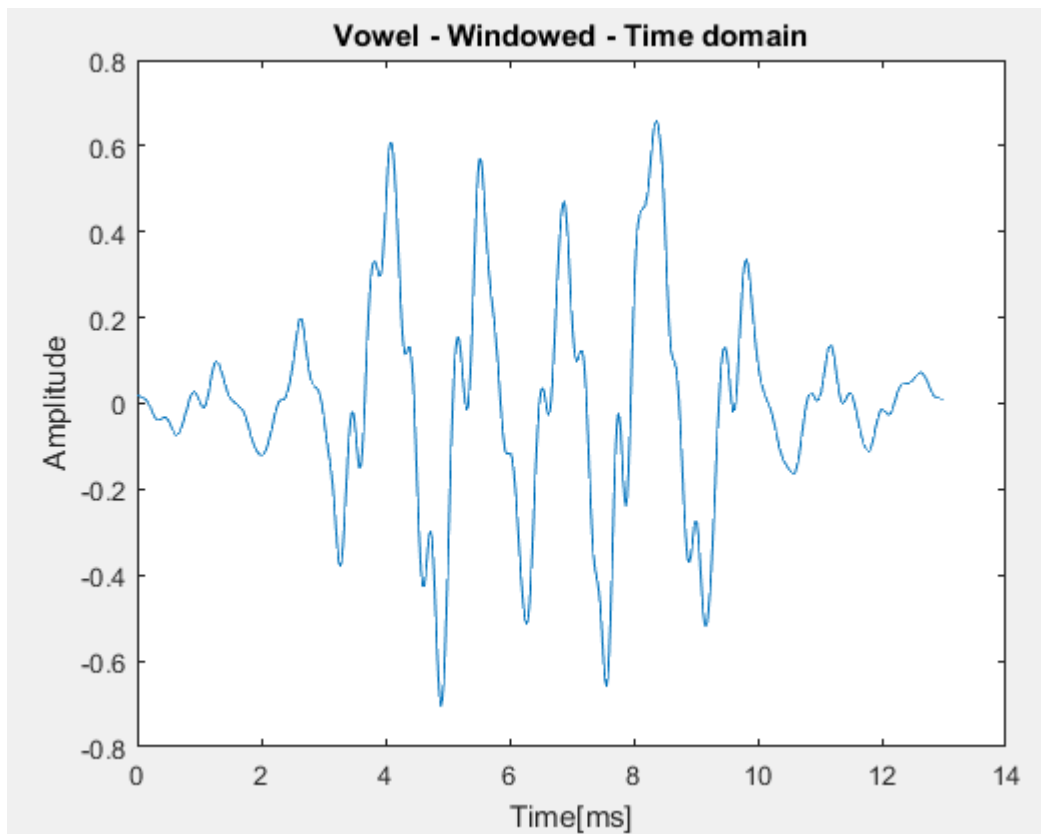
(αρχείο femEEE.m)

Αρχικό σήμα στο χρόνο:

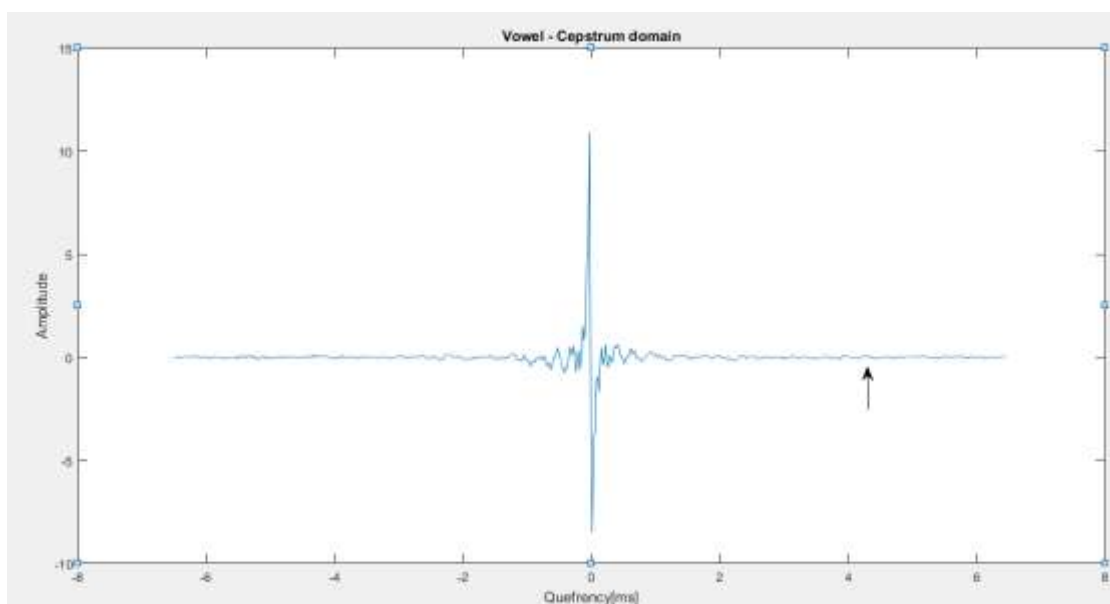


Η εκτιμώμενη θεμελιώδης περίοδος είναι 4.25ms ($\sim 235\text{Hz}$), ενώ τελικά υπολογίζεται στα 4.333ms (230.787Hz).

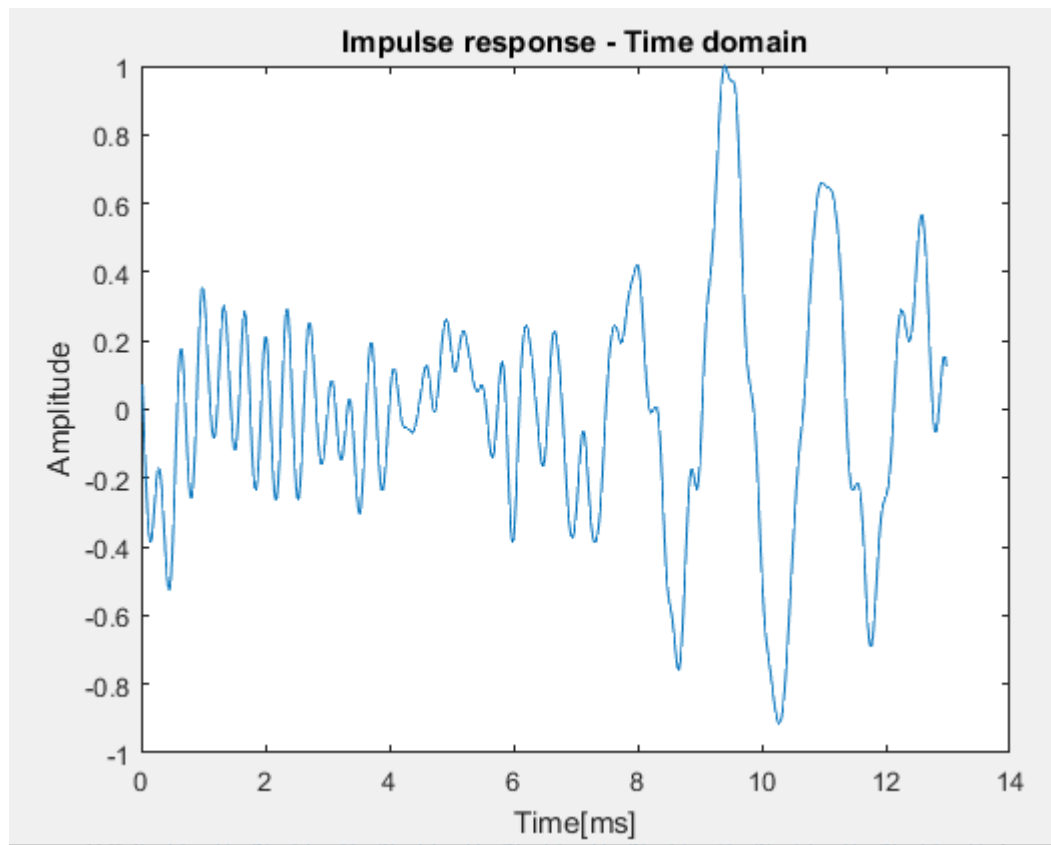
Μετά την εφαρμογή του παραθύρου, το σήμα παίρνει την εξής μορφή:



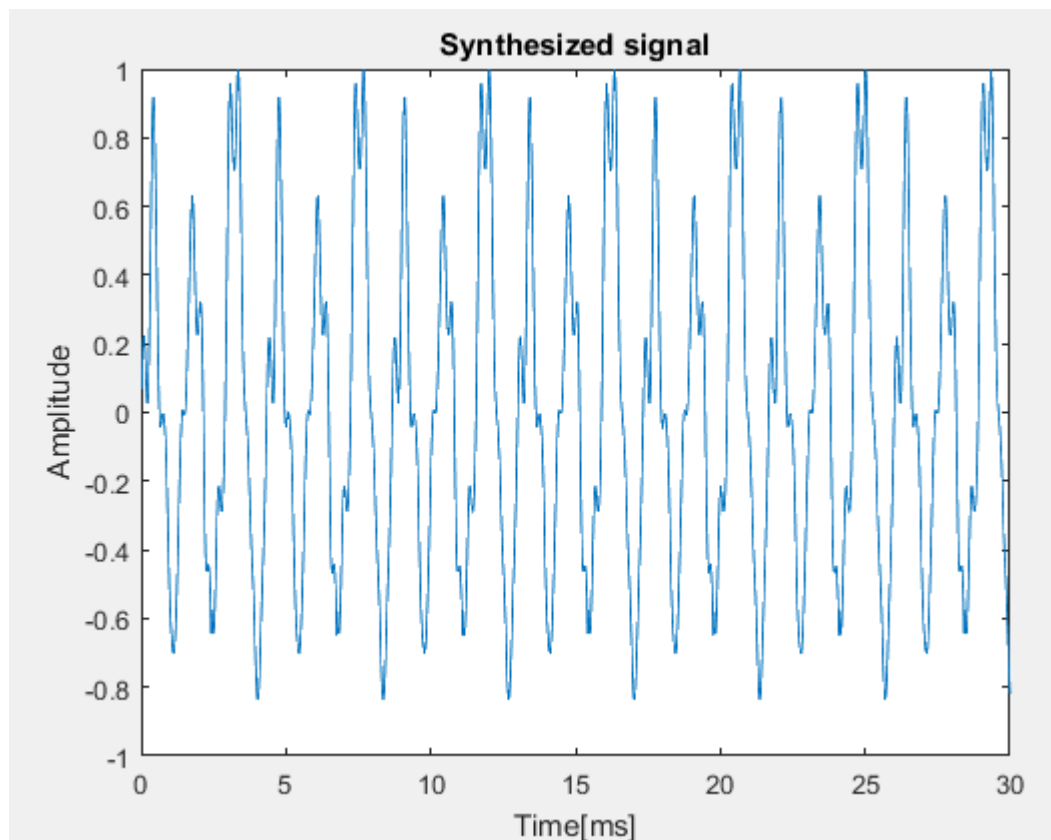
Αναπαριστώντας το σήμα αυτό στο Cepstrum, παρατηρούμε ότι στο σημείο που δείχνει το βέλος υπάρχει κορυφή που υποδεικνύει το pitch.



Μετά την εφαρμογή του χαμηλοπερατού φίλτρου μήκους 0.68 periods προκύπτει η κρουστική απόκριση που αναπαρίσταται στο χρόνο ως εξής:



Τέλος, μετά τη συνέλιξη με τις ώσεις προκύπτει το τελικό σήμα:

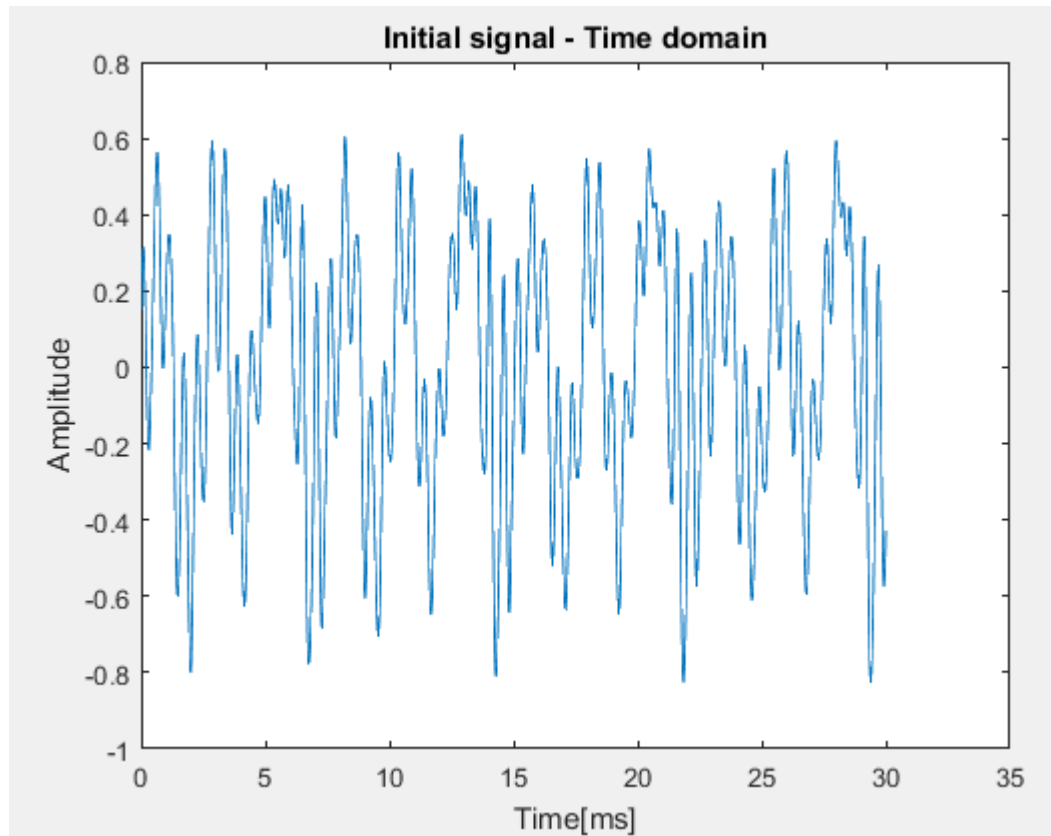


Η θεμελιώδης περίοδος του σήματος αυτού εκτιμάται στα 4.328ms, δηλαδή σχεδόν ίδια με την πραγματική. Επίσης, ακούγοντας τα δυο αρχεία femE.wav και femE_constructed.wav, το δεύτερο θυμίζει το φωνήεν 'έ', ωστόσο δε φέρνει σε ανθρώπινη φωνή.

Ανδρικό 'ε'

(αρχείο maleEEE.m)

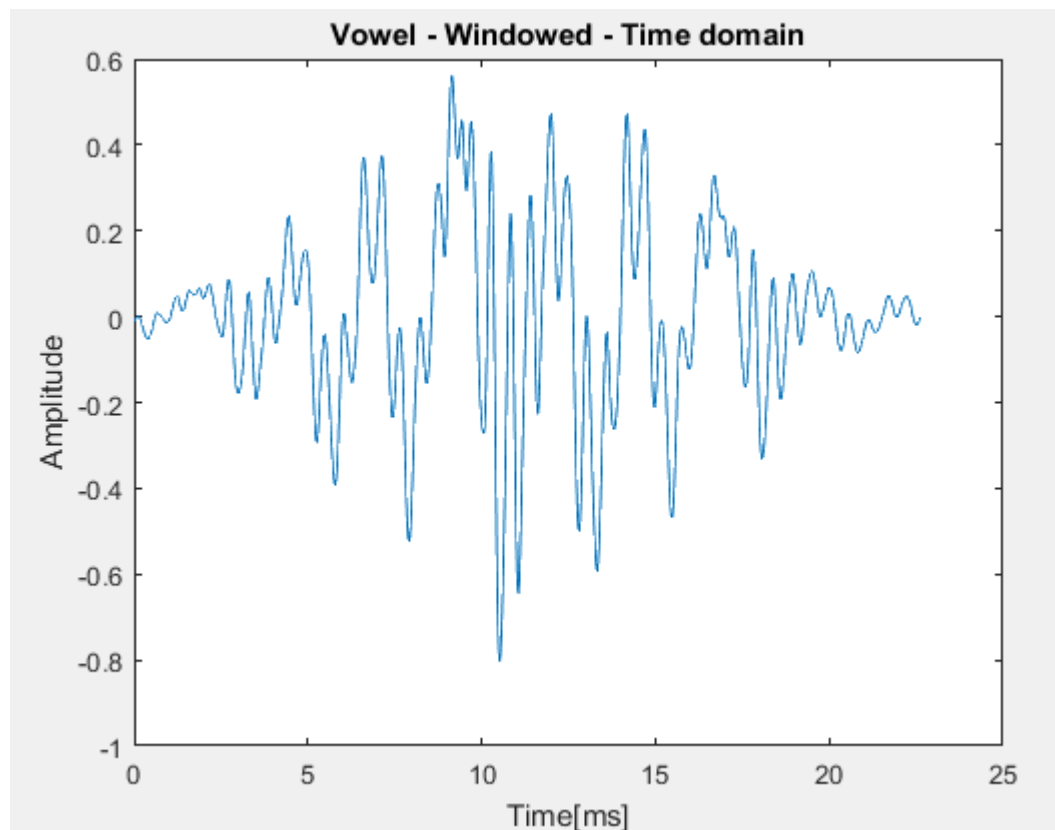
Αρχικό σήμα στο χρόνο:



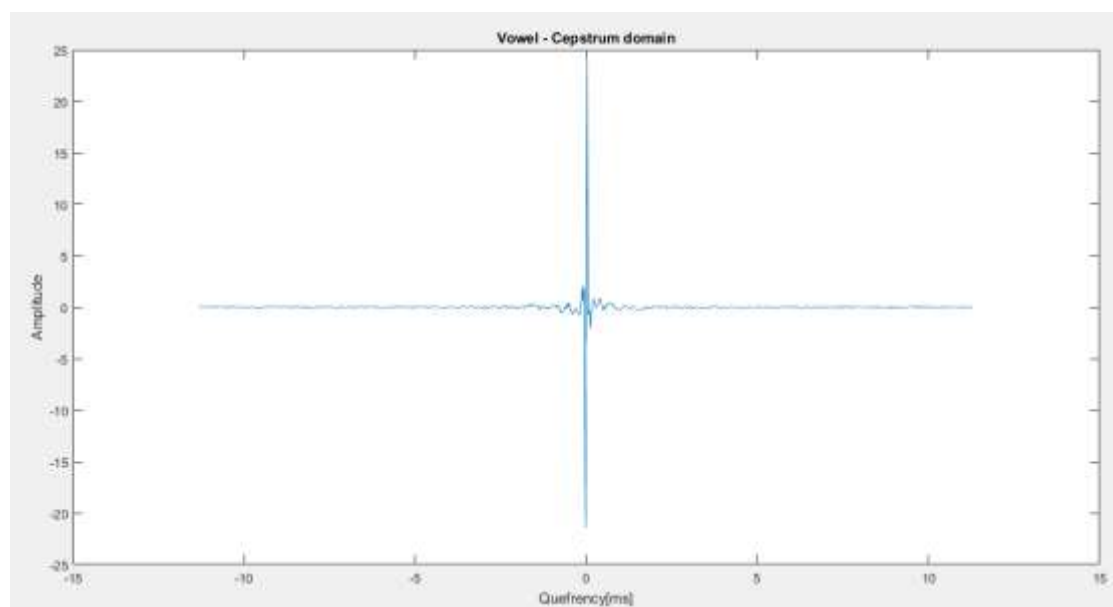
Εκτιμώμενο pitch: 7.58ms ($\sim 132\text{Hz}$)

Ακριβές pitch: 7.5417ms (132.596Hz)

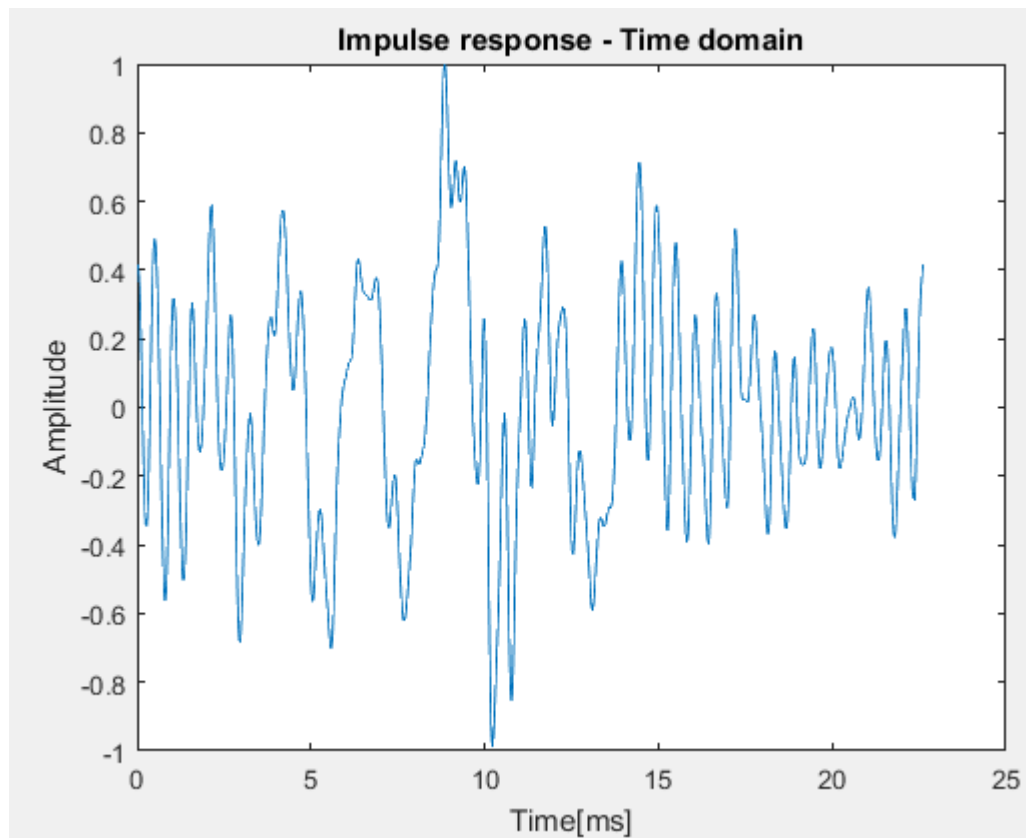
Το σήμα μετά την εφαρμογή Hamming window:



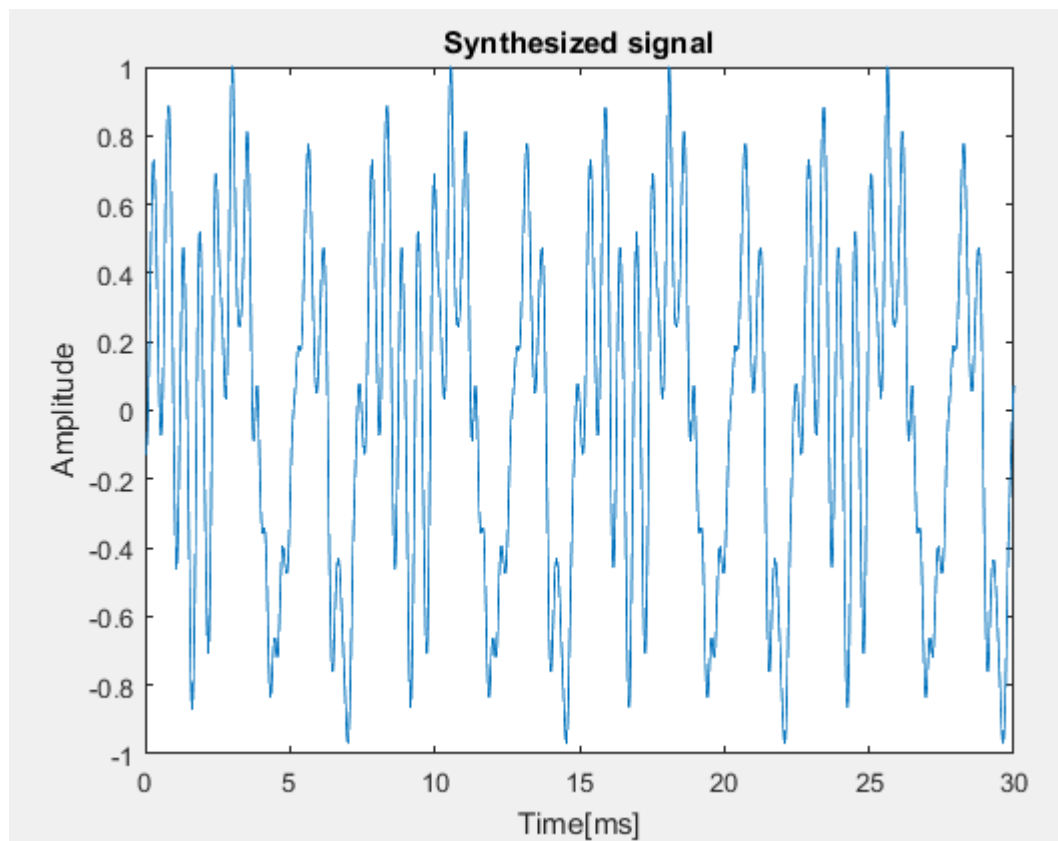
Στον τομέα του Cepstrum, οι κορυφές και πάλι δεν είναι έντονες, και το pitch διακρίνεται μόνο με zoom in στη συγκεκριμένη quefrency.



Μετά την εφαρμογή του χαμηλοπερατού φίλτρου μήκους 0.66 periods προκύπτει η κρουστική απόκριση με κυματομορφή στο χρόνο όπως παρακάτω.



Τέλος, συνελίσσοντας την κρουστική απόκριση με το συρμό ώσεων προκύπτει το τελικό σήμα:



Η περίοδος αυτού του σήματος εκτιμάται στα 7.5ms, δηλαδή σχεδόν όση και το πρωτότυπο.

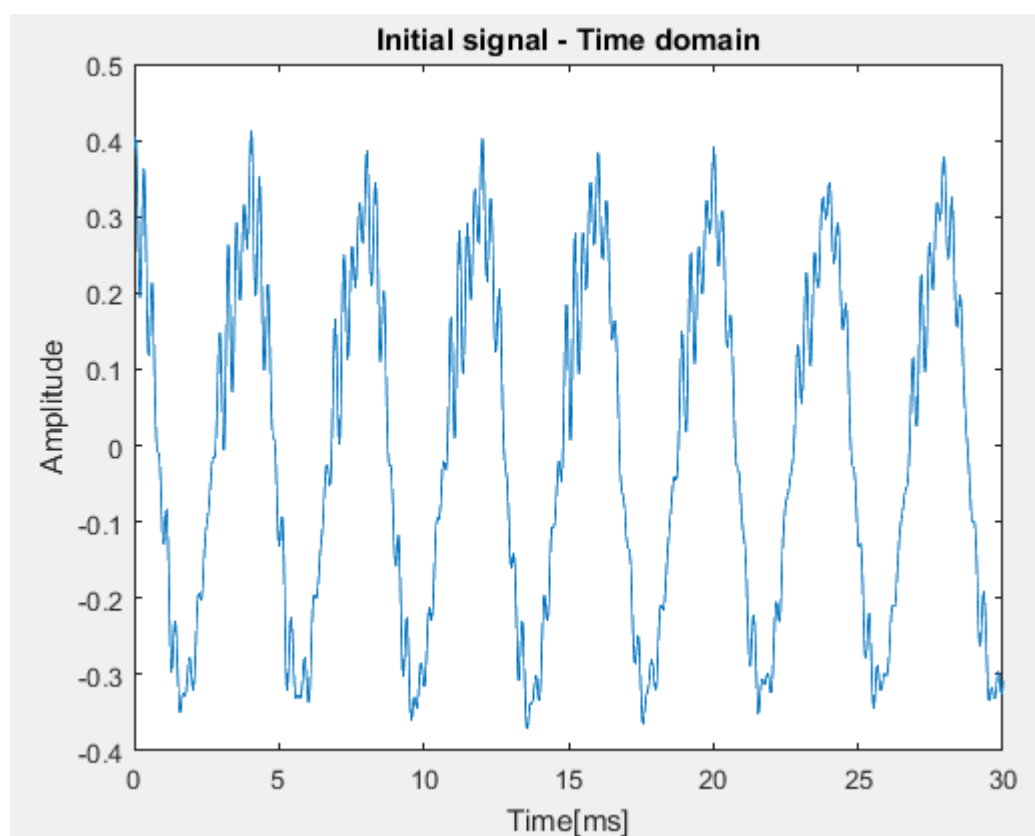
Ακούγοντας τα δυο αρχεία, maleE.wav και maleE_constructed.wav, παρατηρείται ότι το δεύτερο θυμίζει λίγο το φωνήεν 'ε', ωστόσο δε φέρνει ιδιαίτερα σε ανθρώπινη φωνή.

Φωνήεν 'ι'

Γυναικείο 'ι'

(αρχείο femIII.m)

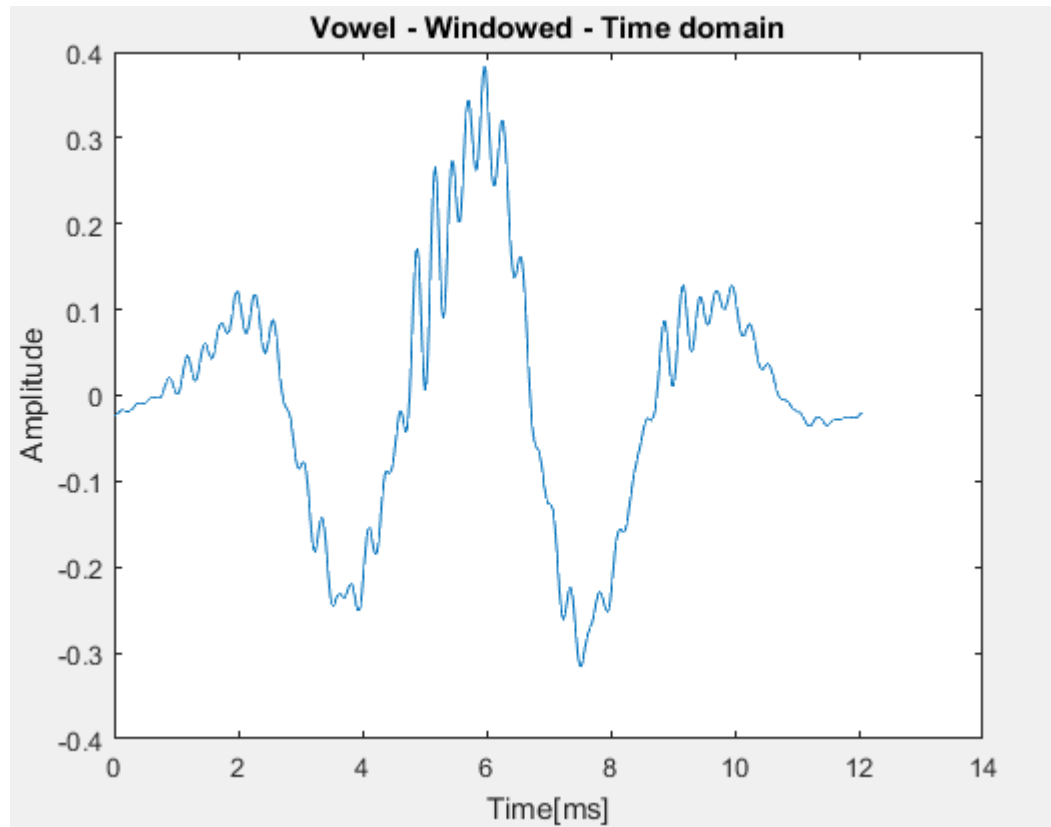
Αρχικό σήμα στο χρόνο:



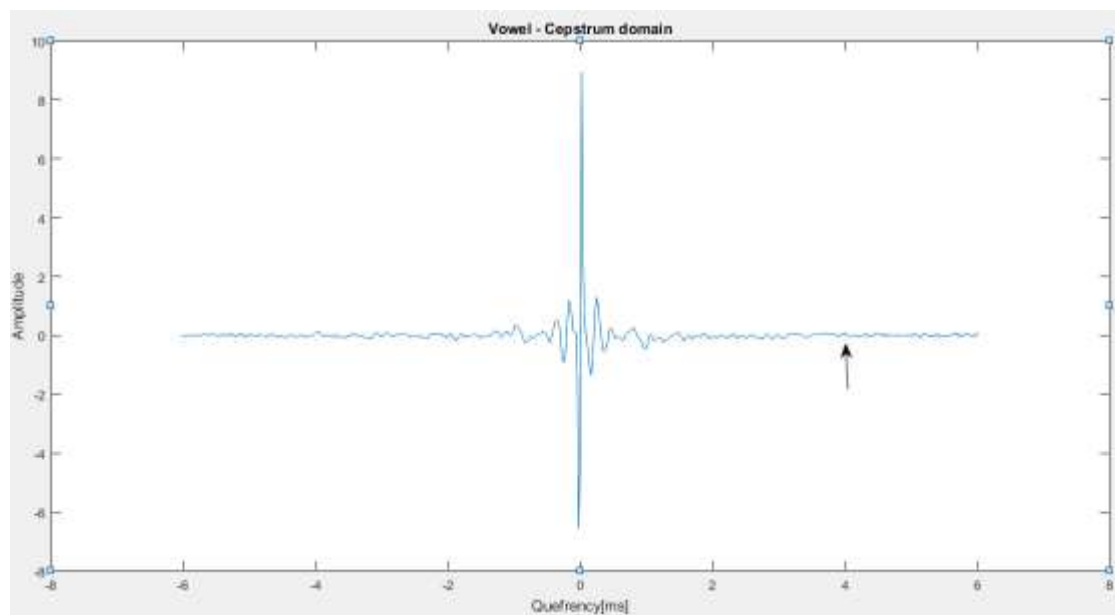
Εκτιμώμενο pitch: 3.978 ms (~251 Hz)

Pitch που προκύπτει από τον τομέα του Cepstrum: 4.0208ms (248.707 Hz)

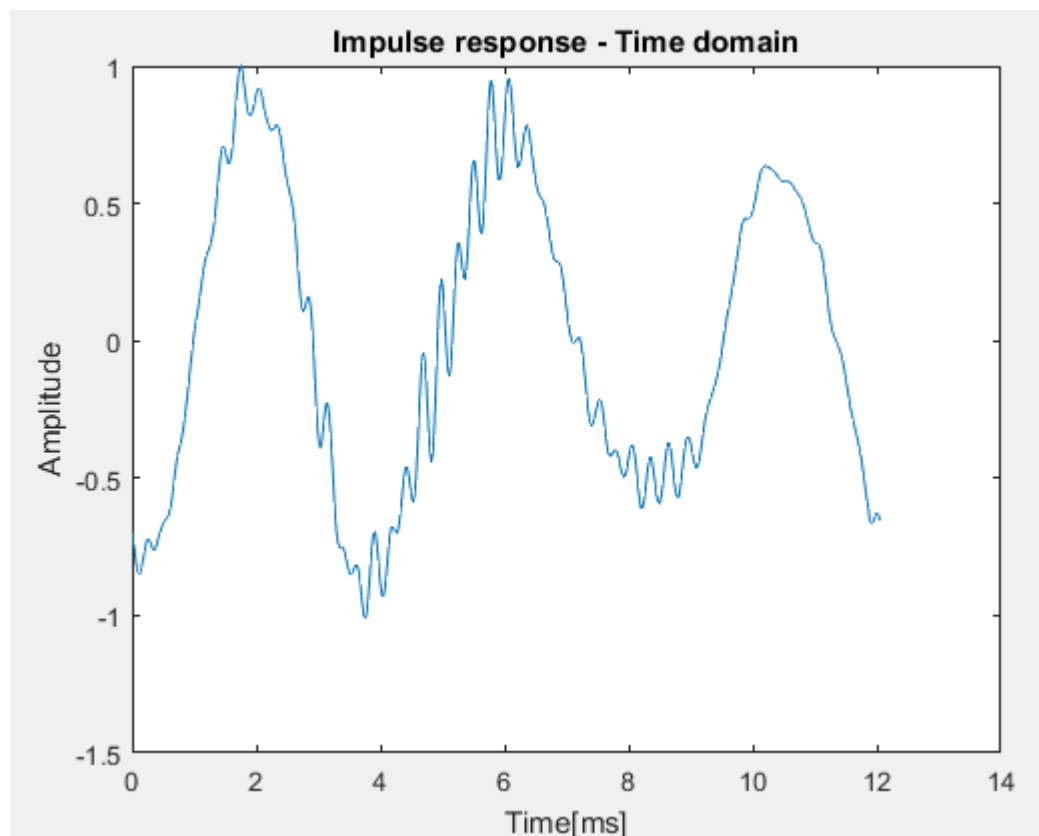
Σήμα που προκύπτει μετά την εφαρμογή του Hamming window:



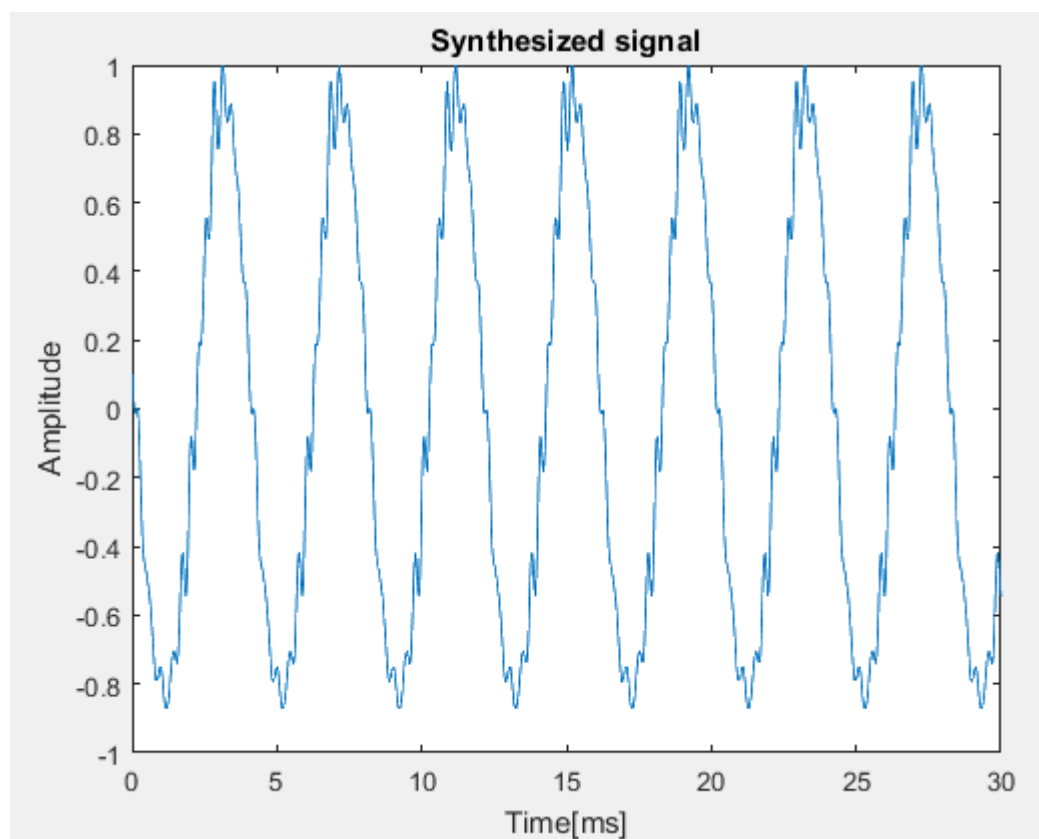
Η απεικόνιση του σήματος στον τομέα του Cepstrum (το βέλος υποδεικνύει το pitch):



Κρουστική απόκριση μετά την εφαρμογή LPF 0.77 periods και την αποσυνέλιξη:



Τελικό κατασκευασμένο σήμα:

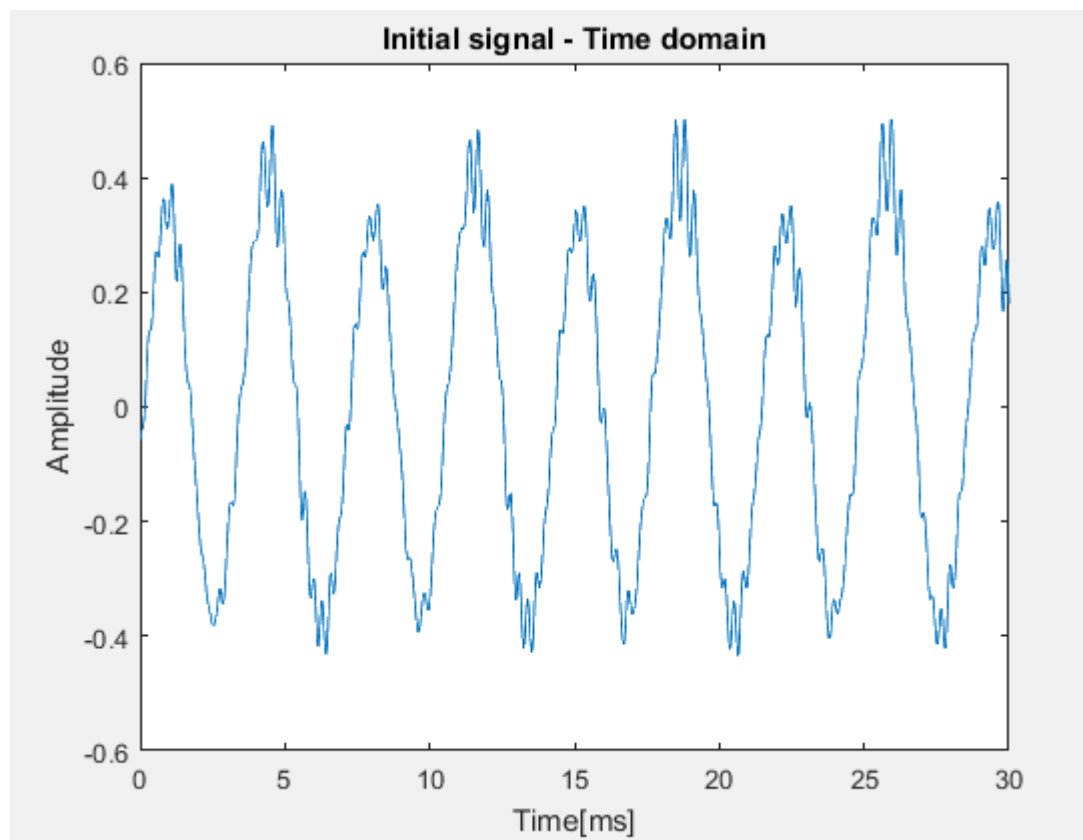


Το σήμα που προκύπτει φαίνεται να έχει περιοδικότητα περίπου 4.022ms (σχεδόν όση και το αρχικό), αλλά και λιγότερες κορυφές σε σχέση με το αρχικό. Ακούγοντας τα αρχεία ήχου, το κατασκευασμένο θυμίζει αρκετά το φωνήεν 'ι'.

Ανδρικό 'ι'

(αρχείο maleIII.m)

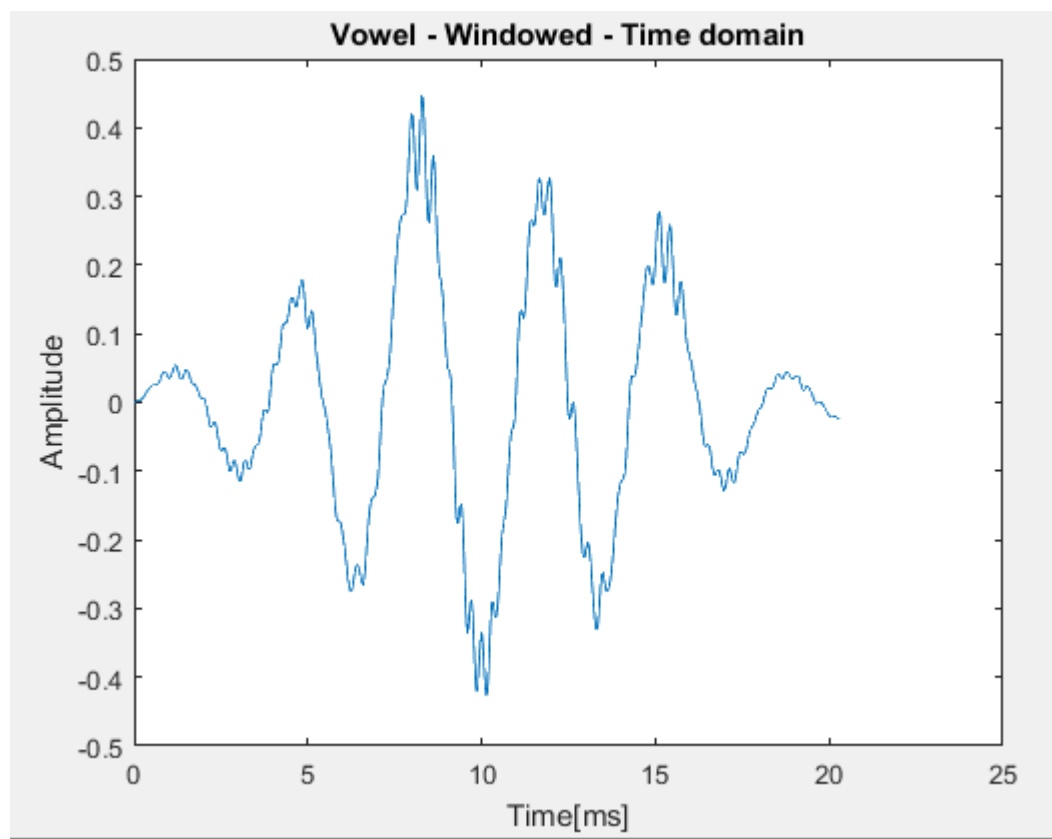
Αρχικό σήμα στο χρόνο:



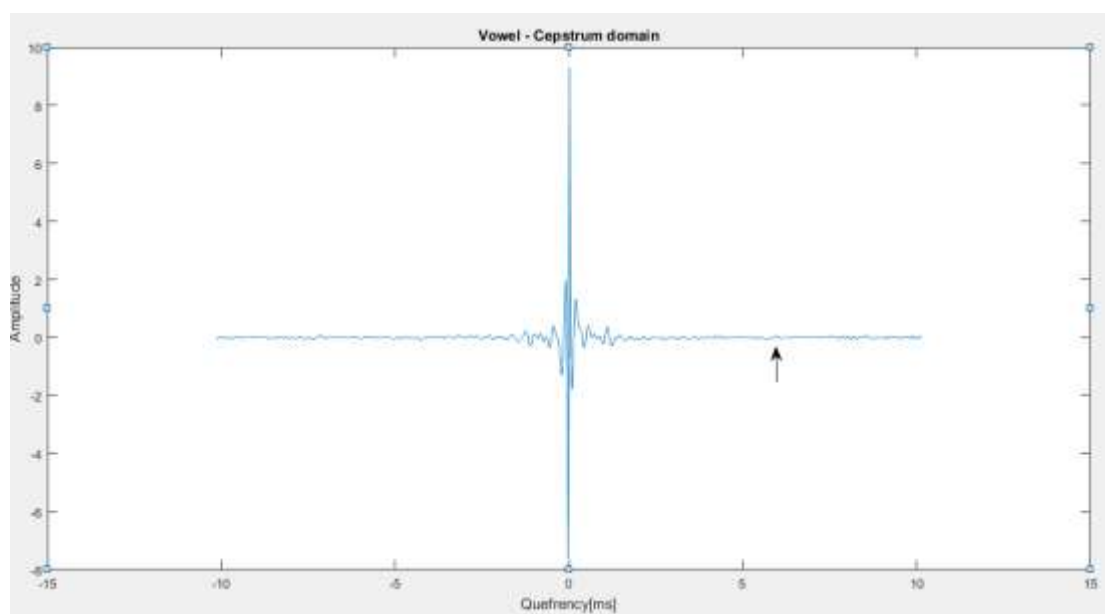
Εκτιμώμενο pitch: 7.108ms (~141 Hz)

Pitch που προκύπτει από τον τομέα του Cepstrum: 6.7708ms (147.693 Hz)

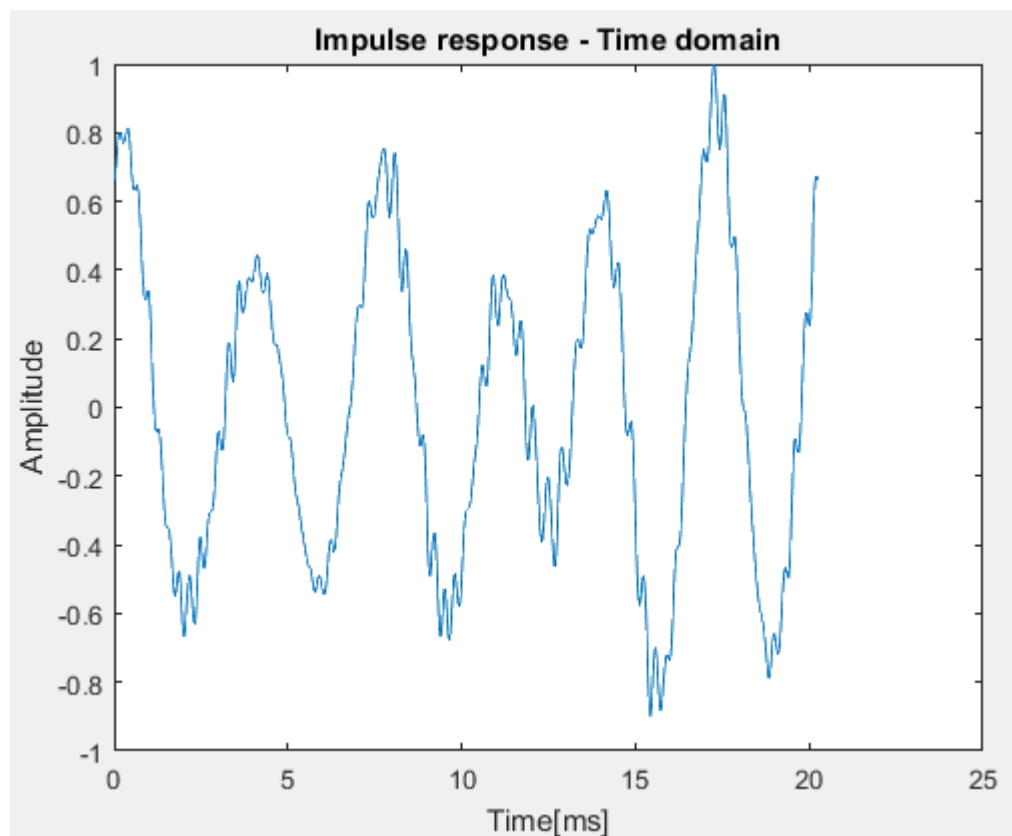
Το σήμα μετά την παραθύρωση:



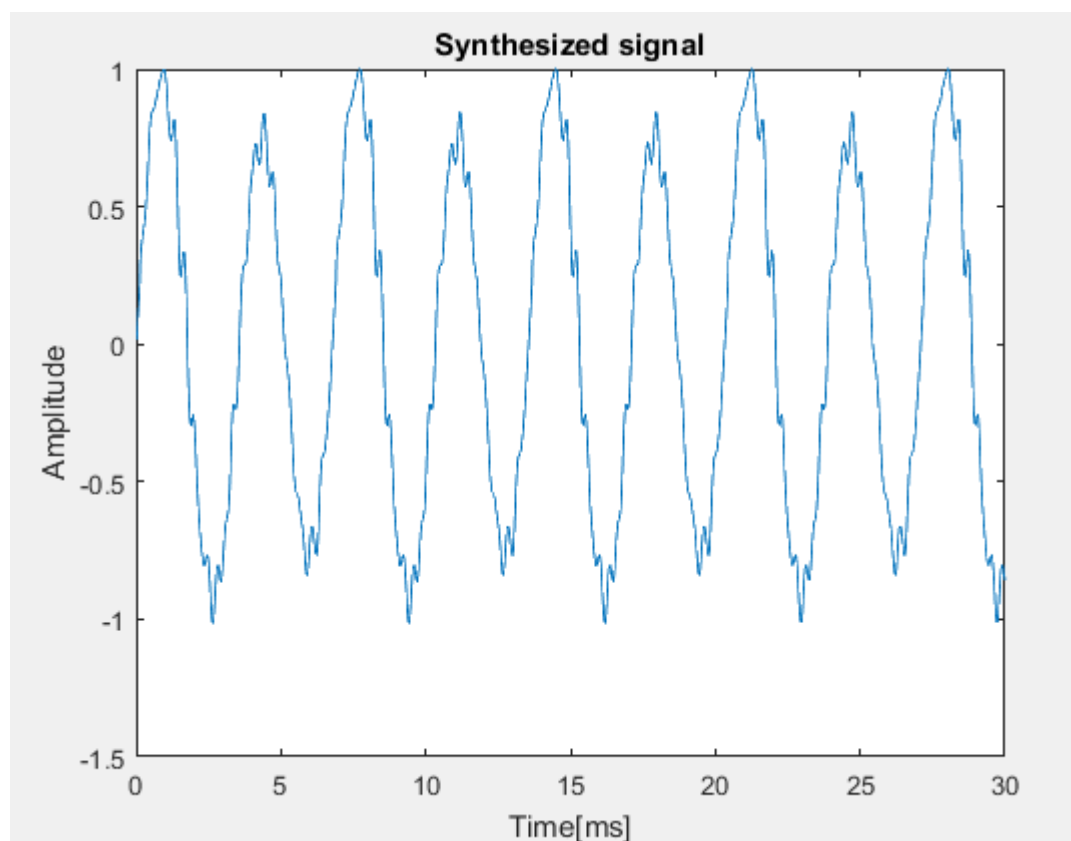
Αναπαράσταση στον τομέα του Cepstrum:



Κρουστική απόκριση (LPF: 0.76 periods):



Τελικό σήμα:



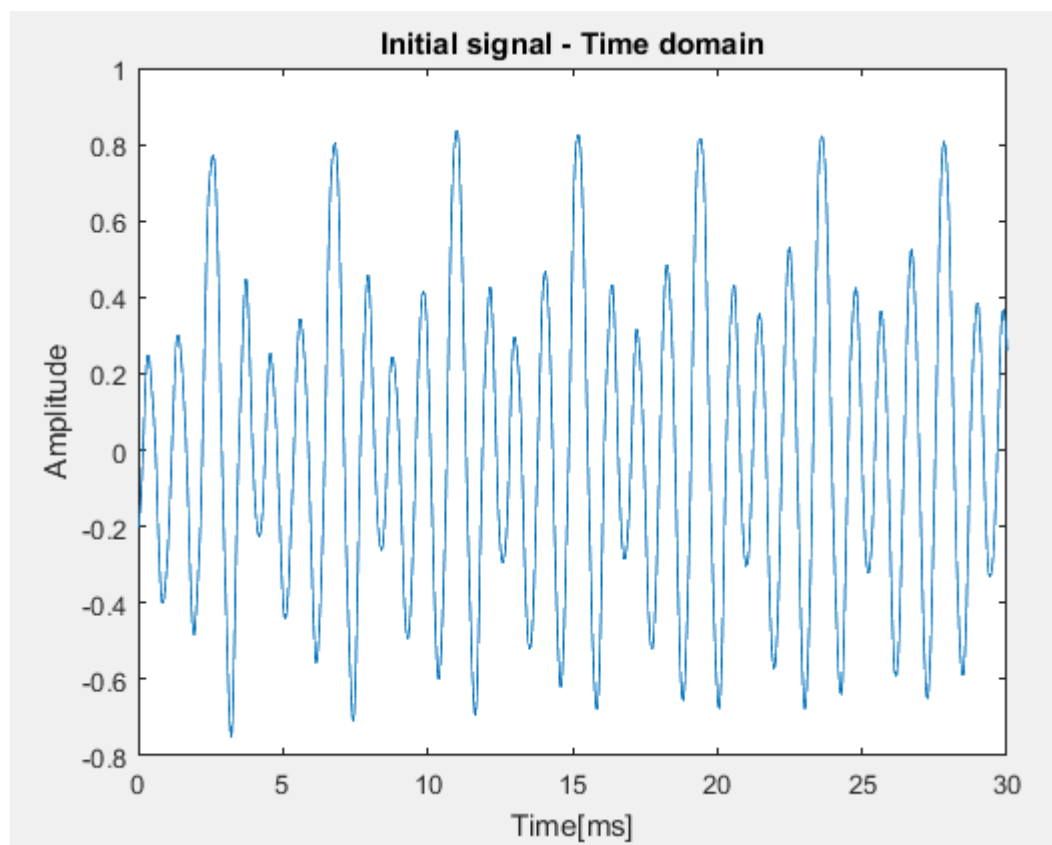
Το σήμα που κατασκευάστηκε δείχνει να έχει σχεδόν ίδια περιοδικότητα με το αρχικό, ωστόσο έχει λιγότερες κορυφές. Ακούγοντάς το, θυμίζει το φωνήεν 'ί', χωρίς να φέρνει ιδιαίτερα σε ανθρώπινη φωνή.

Φωνήεν 'ο'

Γυναικείο 'ο'

(αρχείο fem000.m)

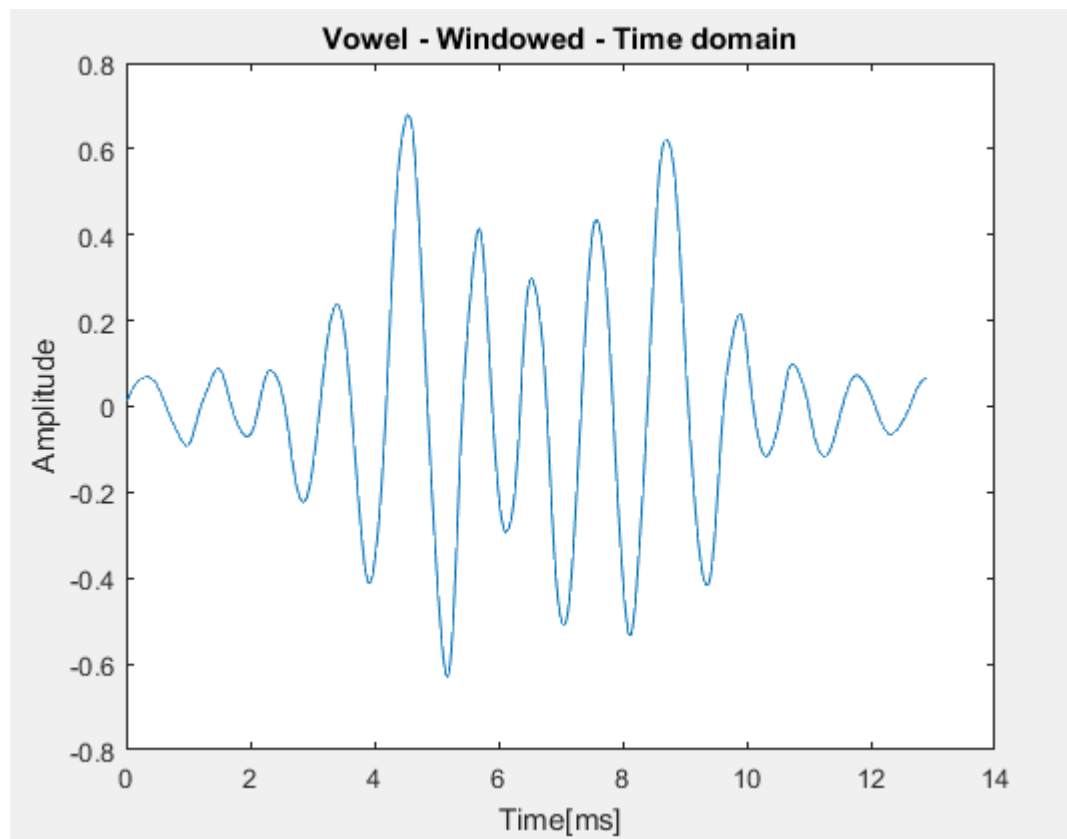
Αρχικό σήμα:



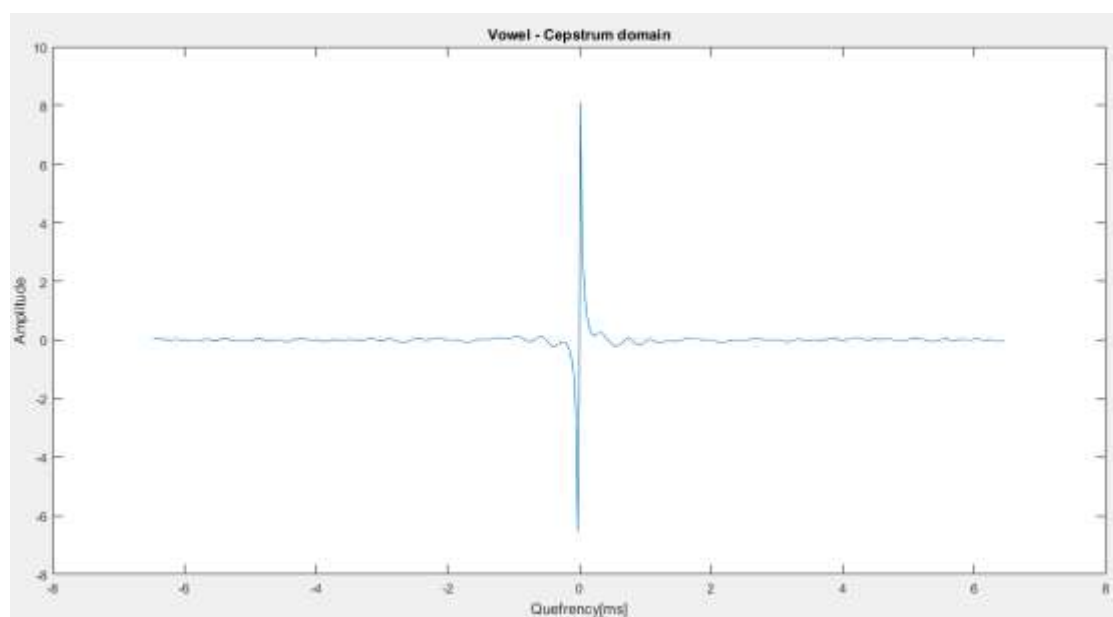
Εκτιμώμενο pitch: 4.208ms (~238 Hz)

Pitch υπολογισμένο από τον Cepstrum τομέα: 4.3125ms (231.884 Hz)

Μετά την εφαρμογή του Hamming Window:

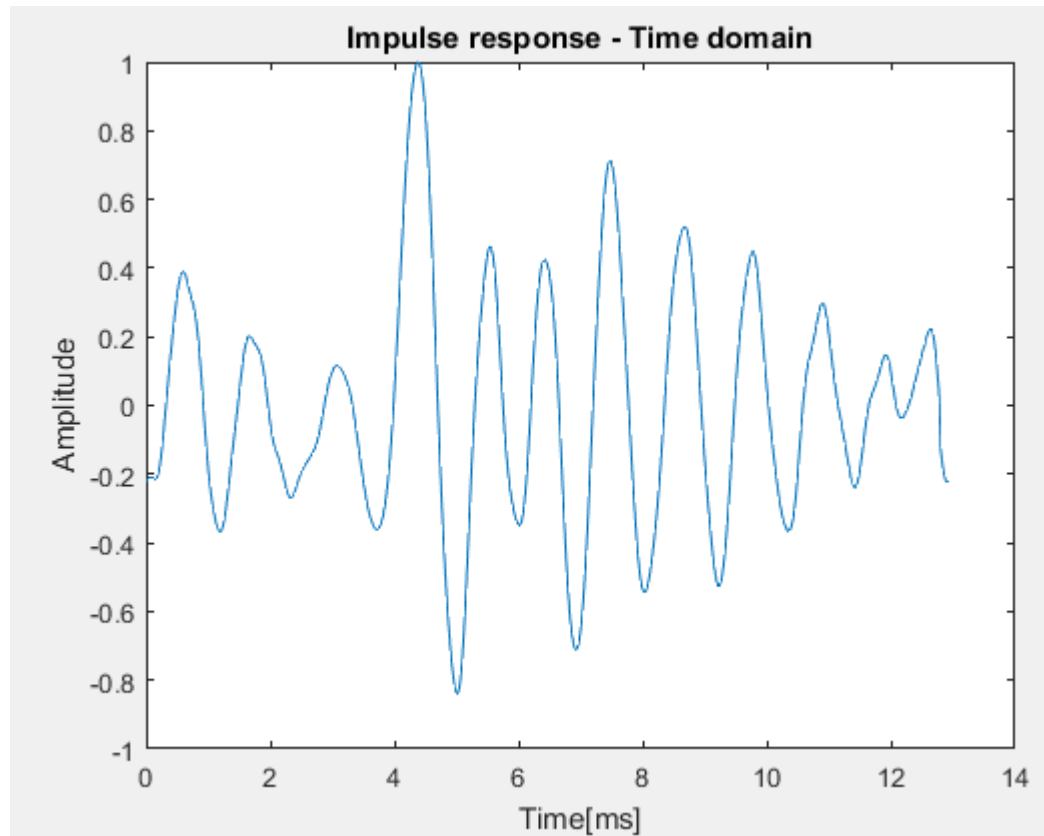


Αναπαράσταση τον τομέα του Cepstrum:



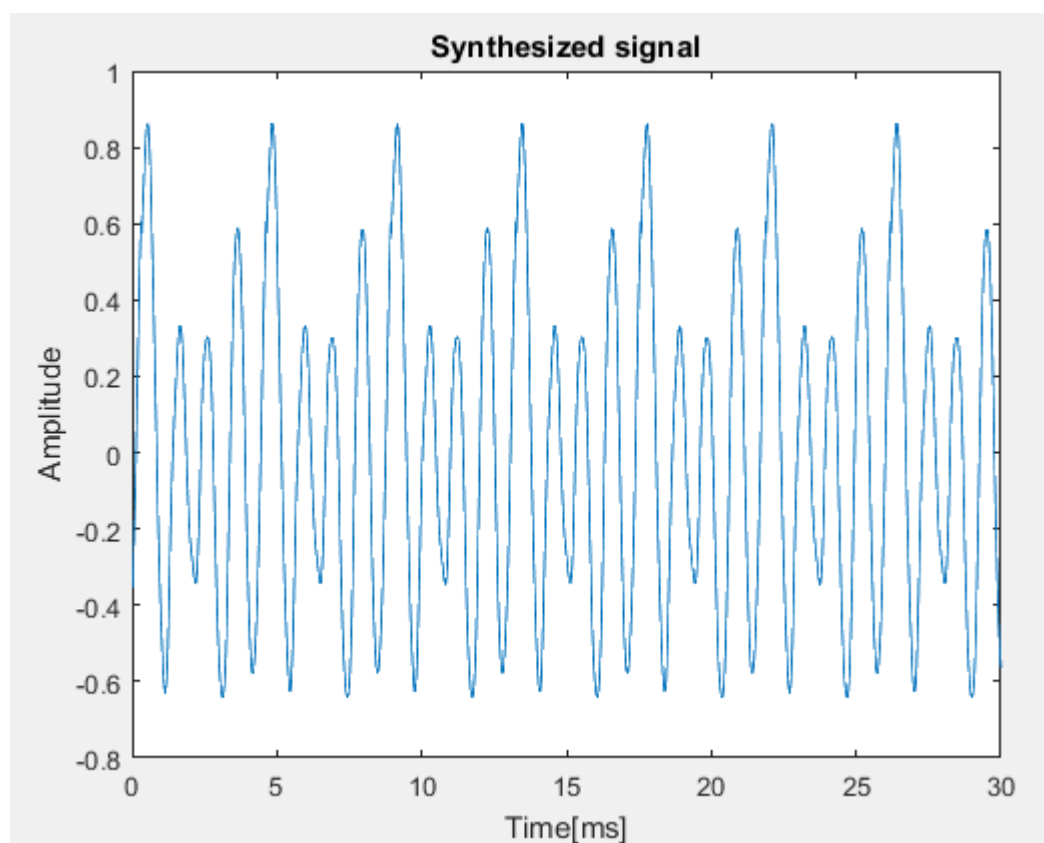
Εδώ είναι φανερό ότι η κυματομορφή έχει πολύ λιγότερες κορυφές σε σχέση τα άλλα θηλυκά φωνήεντα.

Η κρουστική απόκριση μετά την εφαρμογή LPF μήκους 0.8 periods και της αποσυνέλιξης:



Παρατηρούμε ότι, παρά τις διαφορές στο πλάτος, η κρουστική απόκριση είναι αρκετά κοντά στο σήμα που προκύπτει από την εφαρμογή του παραθύρου.

Τέλος, το σήμα που κατασκευάζεται:

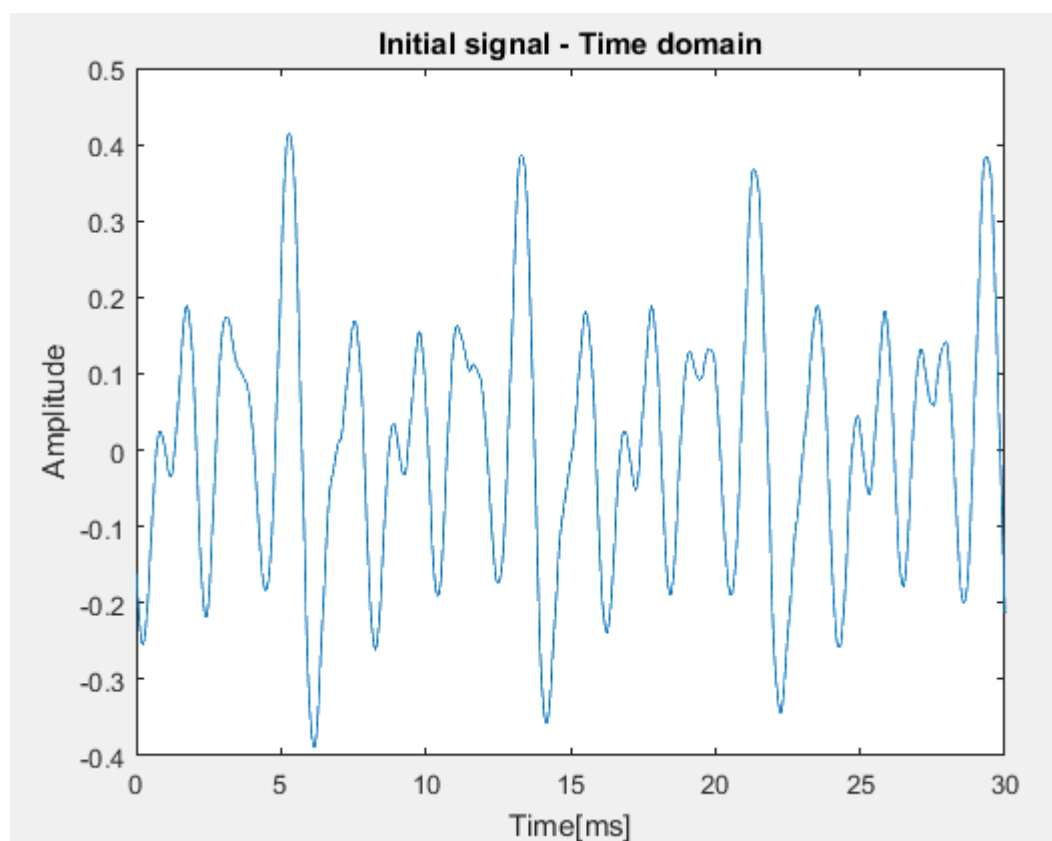


Το σήμα αυτό, παρόλο που οπτικά φαίνεται να είναι κοντά στο πρωτότυπο, ακούγοντας το αρχείο ήχου, θυμίζει λίγο το φωνήεν 'ο' και μάλλον καθόλου ανθρώπινη φωνή.

Ανδρικό 'ο'

(αρχείο male000.m)

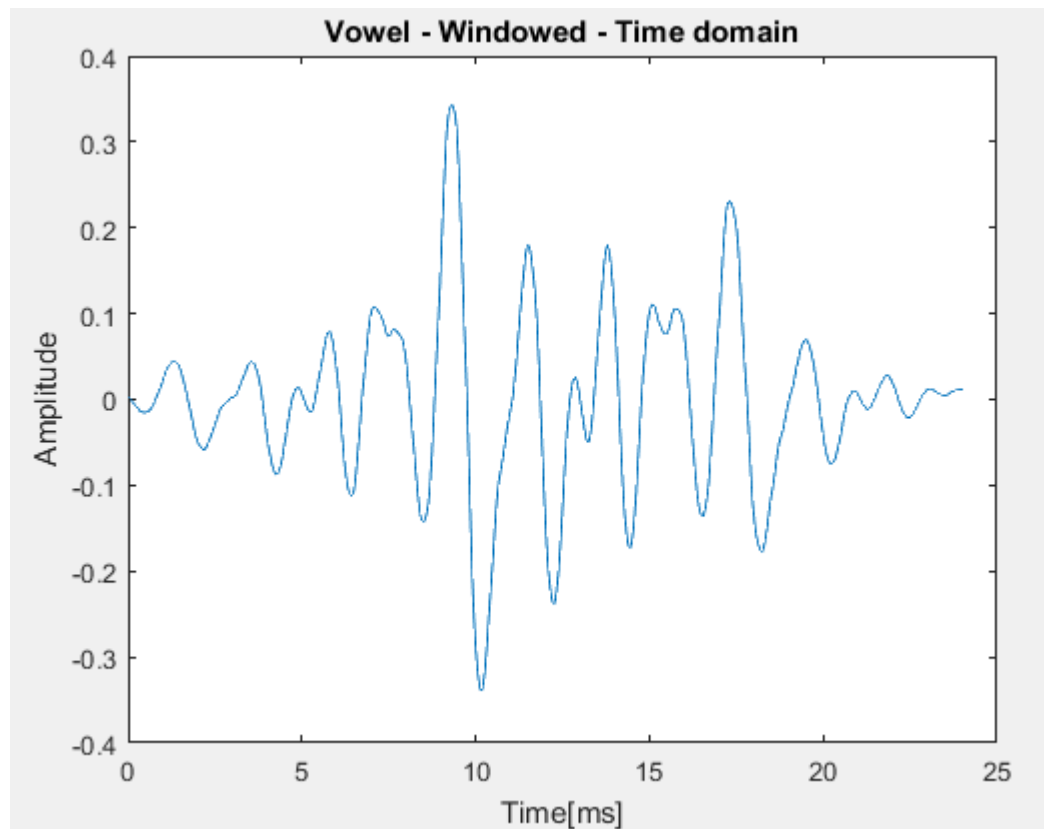
Αρχικό σήμα:



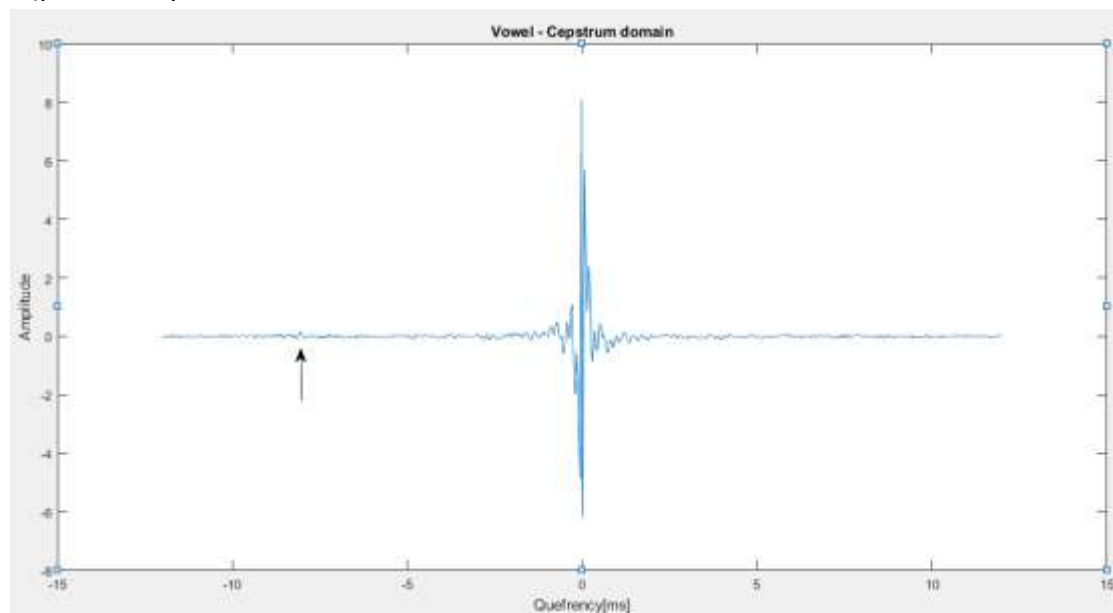
Εκτιμώμενο pitch: 8.04ms (~124 Hz)

Pitch υπολογισμένο από τον τομέα του Cepstrum: 8.0208ms (124.676 Hz)

Μετά την εφαρμογή του Hamming window:

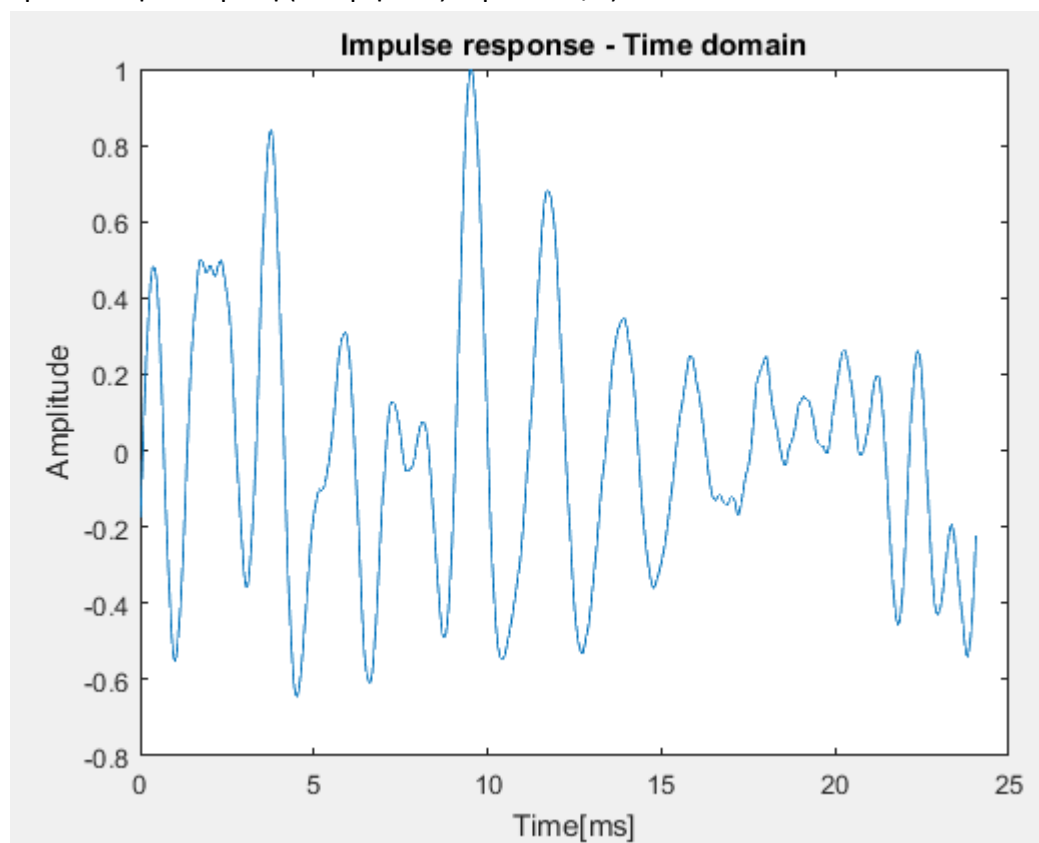


Σήμα στο Cepstrum:

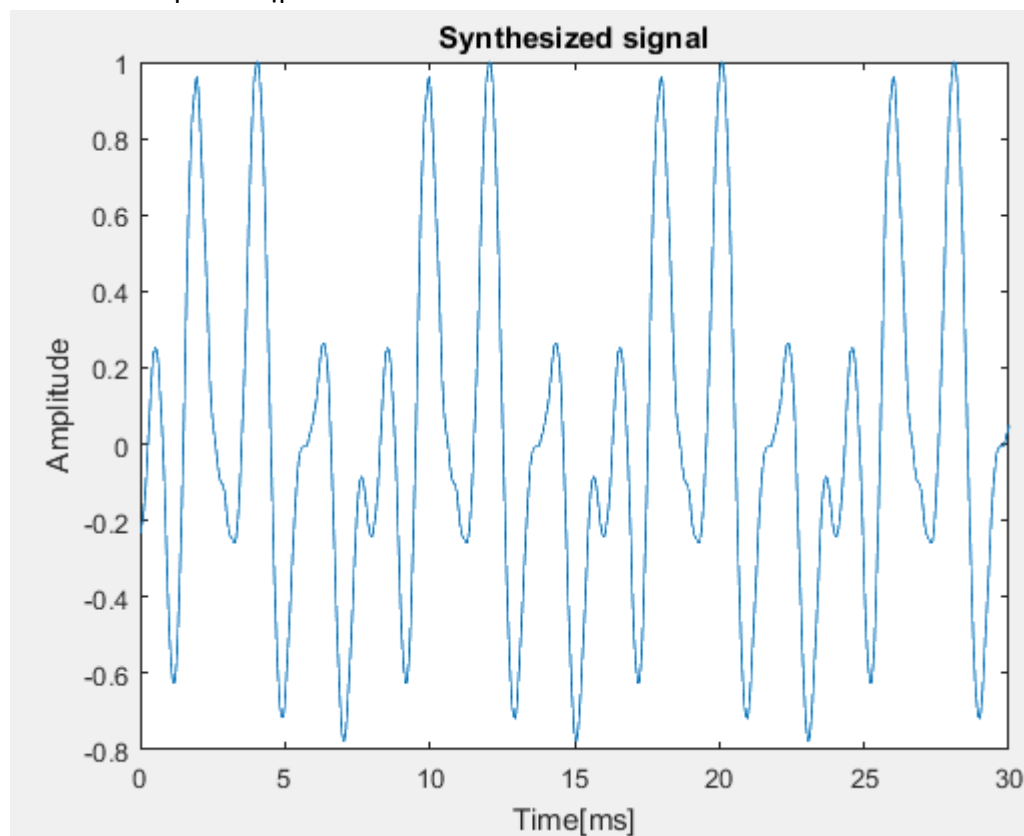


Στον αρνητικό άξονα το pitch φαίνεται να είναι πιο ξεκάθαρο.

Κρουστική απόκριση (LPF μήκους $2 \cdot \text{periods}/3$):



Κατασκευασμένο σήμα:



Το σήμα οπτικά φαίνεται να έχει μια περιοδικότητα ίδια με το αρχικό (περίπου στα 8ms) και ακούγοντας το αντίστοιχο αρχείο ήχου δείχνει να είναι μια αρκετά καλή

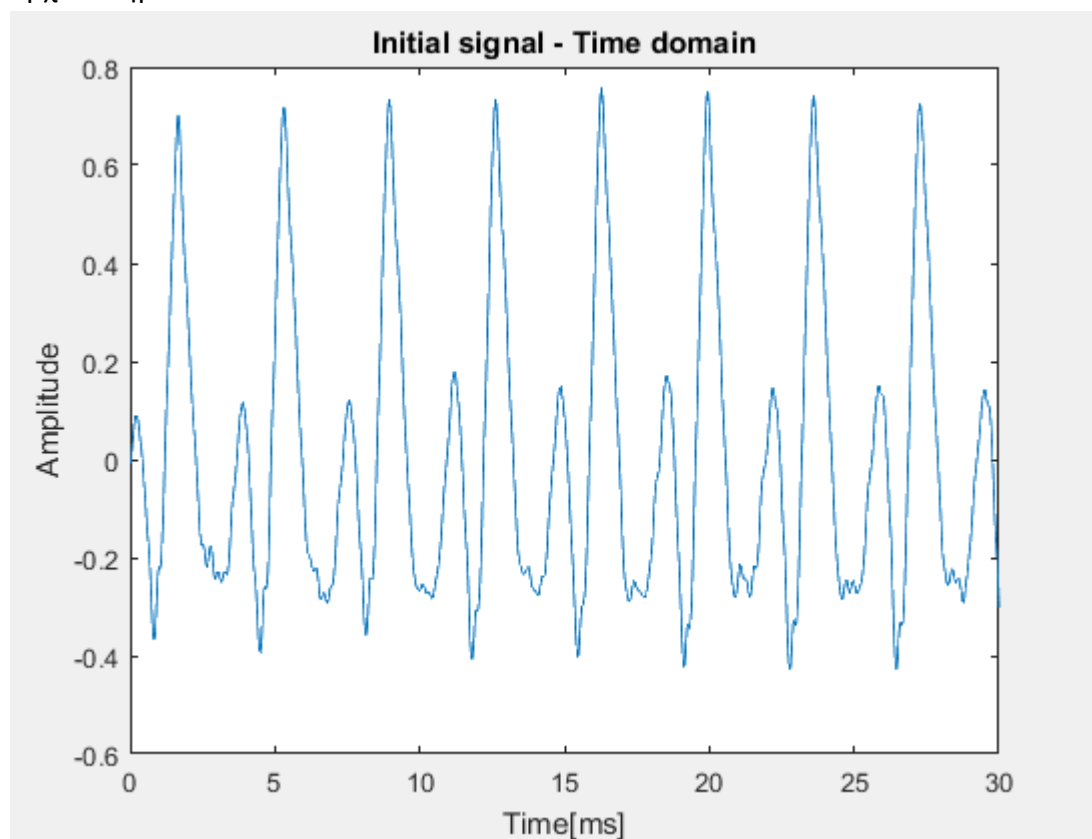
προσέγγιση του πρωτότυπου, καθώς θυμίζει το φωνήεν 'ο', ενώ φέρνει και λίγο σε ανθρώπινη φωνή.

Φωνήεν 'ου'

Γυναικείο 'ου'

(αρχείο femUUU.m)

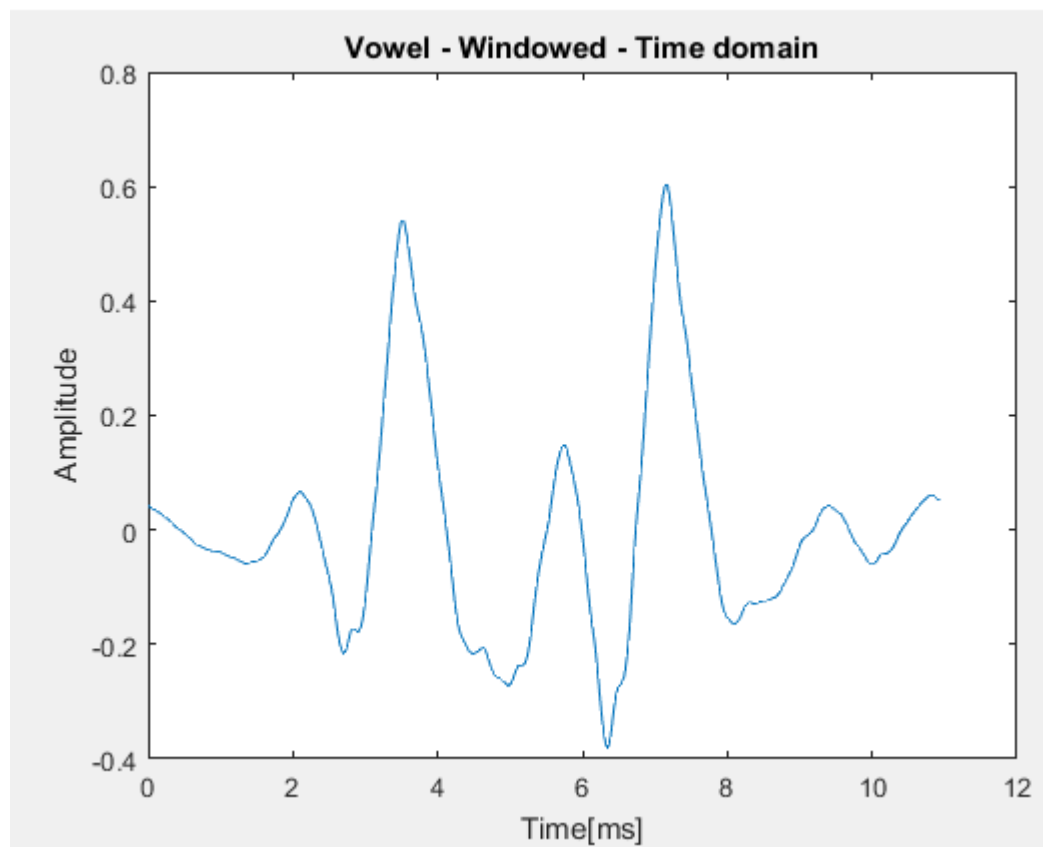
Αρχικό σήμα:



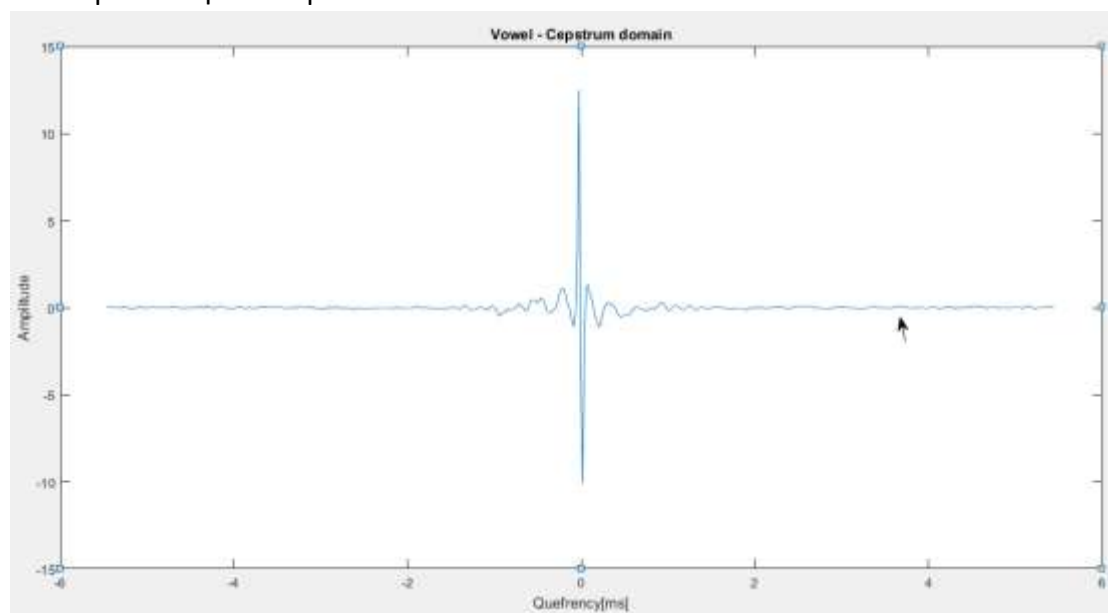
Εκτιμώμενο Pitch: 3.646ms (~ 274 Hz)

Pitch υπολογισμένο από τον τομέα του Cepstrum: 3.6458ms (274.288 Hz)

Μετά την εφαρμογή του Hamming window:

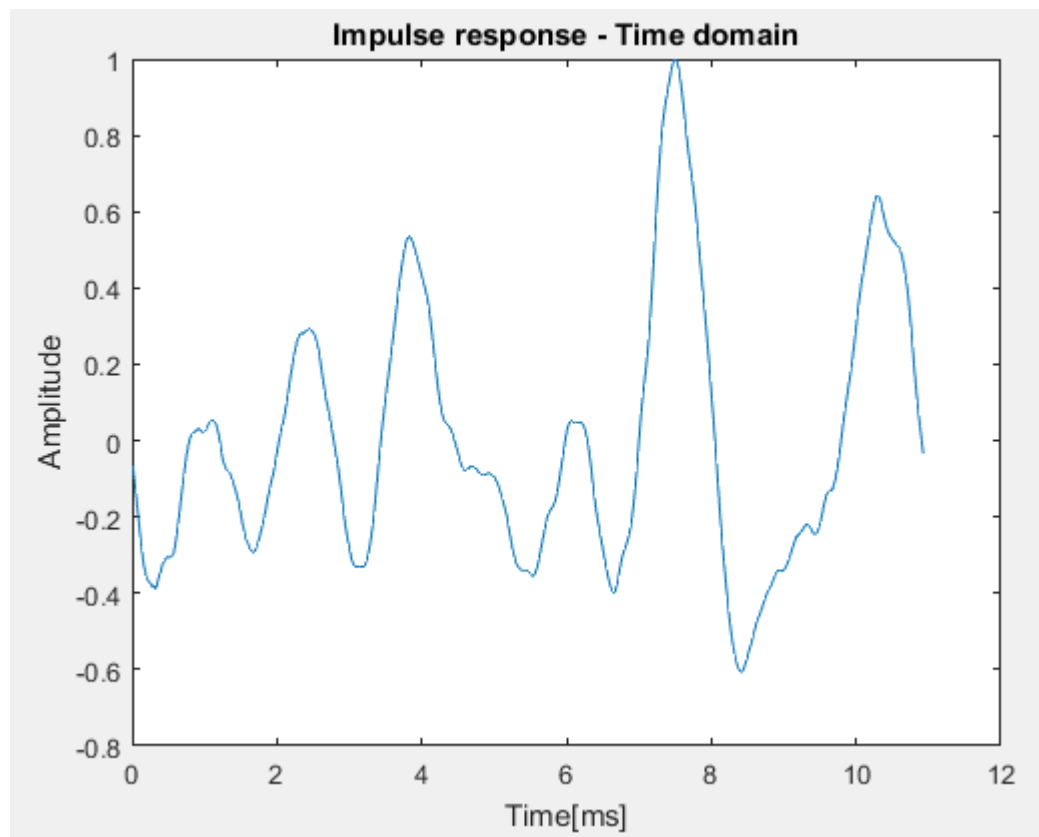


Αναπαράσταση στο Cepstrum:

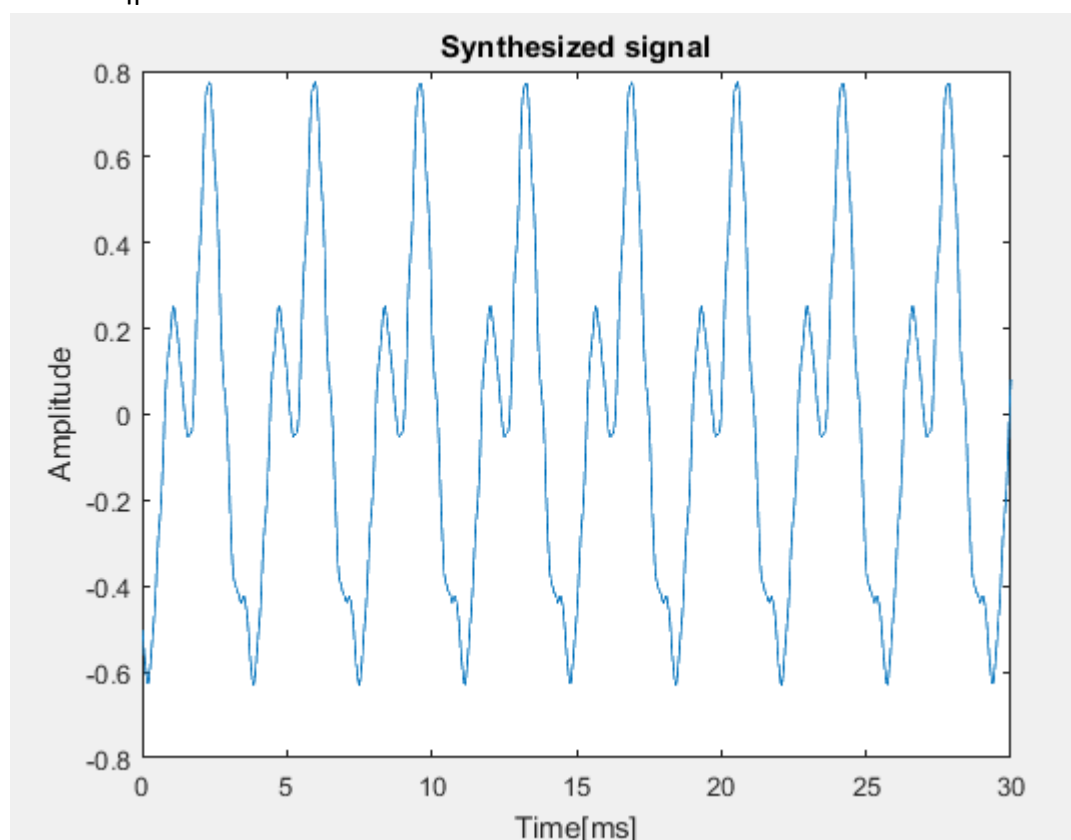


Είναι φανερό ότι χωρίς zoom in το συγκεκριμένο pitch δε διακρίνεται σχεδόν καθόλου.

Κρουστική απόκριση μετά από εφαρμογή LPF μήκους 0.75 periods και αποσυνέλιξη:



Τελικό σήμα:



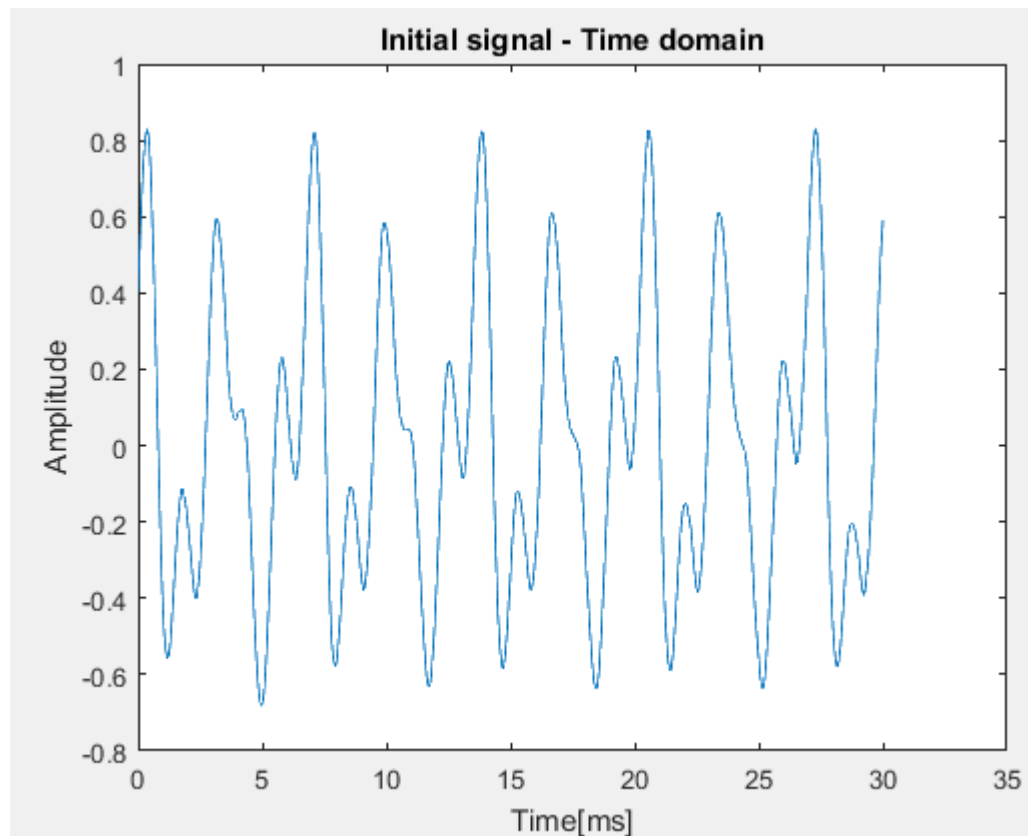
Το σήμα αυτό εκτιμάται να έχει θεμελιώδη περίοδο 3.645ms (δηλαδή όση και το αρχικό).

Ακούγοντας τα δυο αρχεία ήχου το κατασκευασμένο σήμα φαίνεται να είναι μια αρκετά καλή προσέγγιση του αρχικού.

Ανδρικό 'ου'

(αρχείο maleUUU.m)

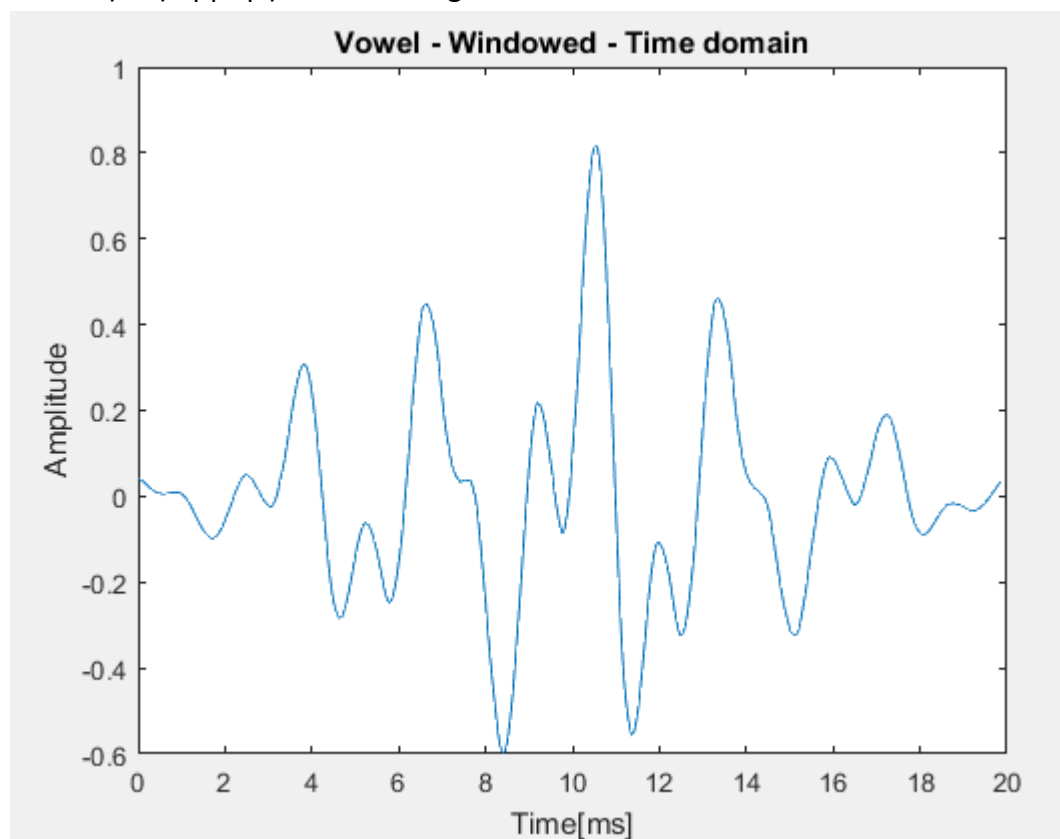
Αρχικό σήμα:



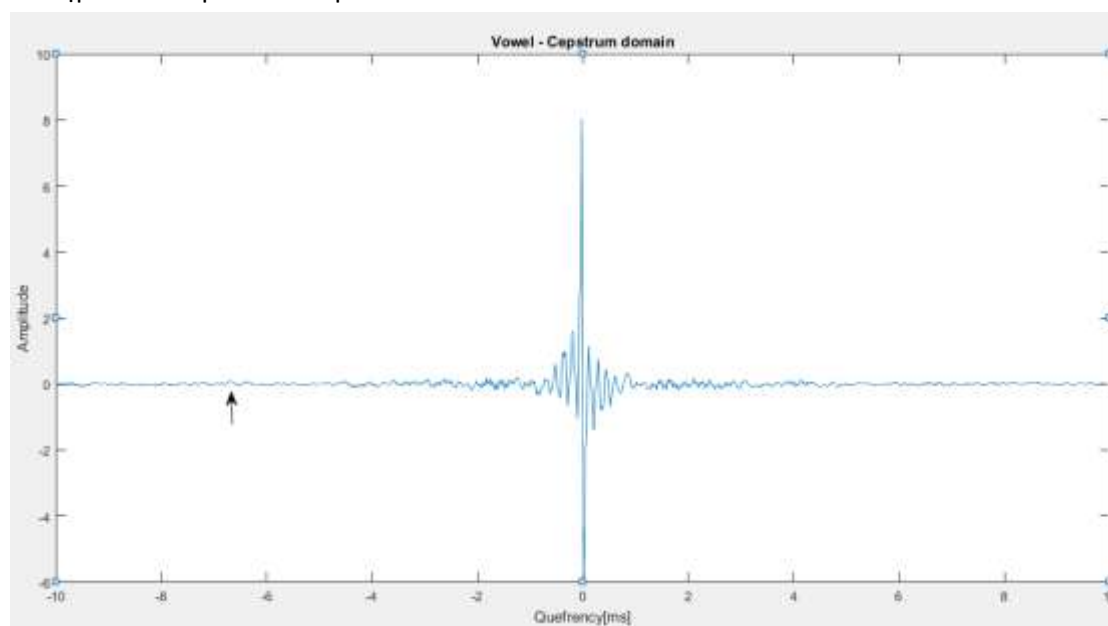
Εκτιμώμενο Pitch: 6.746ms (~ 148 Hz)

Pitch υπολογισμένο από τον τομέα του Cepstrum: 6.625ms (150.943 Hz)

Μετά την εφαρμογή του Hamming window:

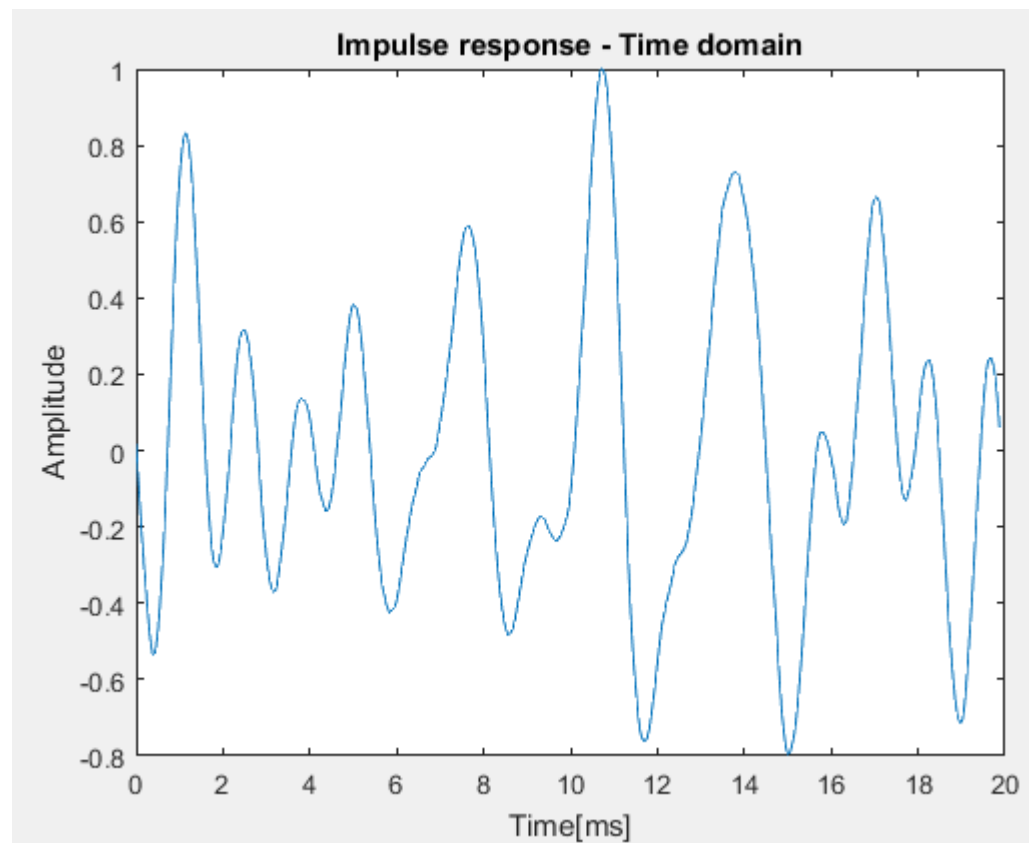


Το σήμα στον τομέα του Cepstrum:

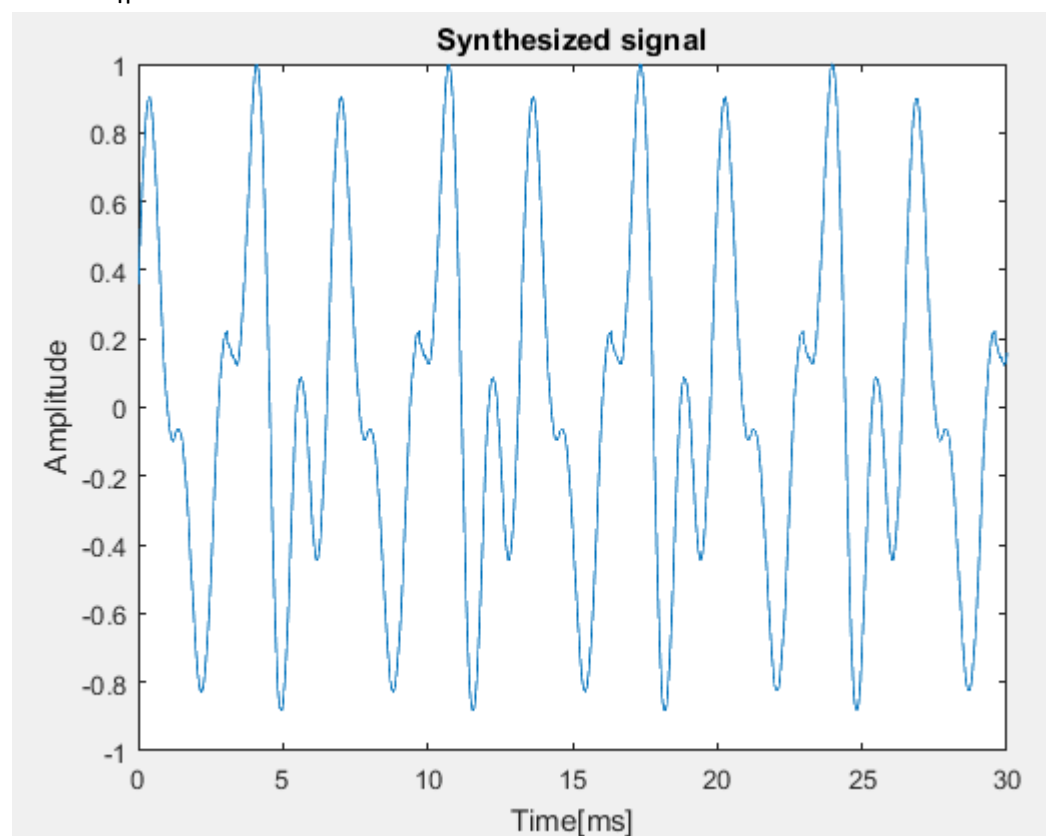


Συγκρίνοντάς το με το αντίστοιχο γυναικείο, το συγκεκριμένο φαίνεται να έχει περισσότερες κορυφές.

Κρουστική απόκριση μετά από εφαρμογή LPF μήκους 0.73 periods και αποσυνέλιξη:



Τελικό σήμα:



Το σήμα οπτικά μοιάζει αρκετά στο πρωτότυπο και ακούγοντας τα αρχεία ήχου παρατηρούμε ότι το κατασκευασμένο θυμίζει το φωνήεν 'ου' και φέρνει λίγο σε ανθρώπινη φωνή.

Παρατηρήσεις - Συμπεράσματα

-Από τις παραπάνω αναλύσεις είναι φανερό ότι σε κάθε φωνήεν το ανδρικό pitch είναι περίπου 100Hz χαμηλότερο από το αντίστοιχο γυναικείο. Στον παρακάτω πίνακα παρουσιάζονται οι θεμελιώδεις συχνότητες που μετρήθηκαν, σε Hz.

	‘α’	‘ε’	‘ι’	‘ο’	‘ου’
Γυναίκα	220	230	249	232	274
Άνδρας	127	133	148	125	151

-Στον τομέα του Cerstrum παρατηρούνται εν γένει περισσότερες κορυφές στα θηλυκά φωνήεντα σε σχέση με τα ανδρικά. Αυτό πιθανόν συμβαίνει επειδή η γυναικεία φωνή έχει πιο “breathy” ποιότητα με αποτέλεσμα στον τομέα του Cerstrum αυτό να μεταφράζεται ως θόρυβος, ενώ αντίθετα το αντρικό Cerstrum να είναι πιο ομαλό.

-Ακούγοντας τα αρχεία ήχου από τα κατασκευασμένα φωνήεντα παρατηρούμε ότι είναι αρκετά μειωμένο το “buzziness”. Αυτό οφείλεται στο γεγονός ότι η κρουστική απόκριση που υπολογίζεται είναι μικτής φάσης. Παρ’ όλα αυτά, στα περισσότερα αρχεία αν και ο ήχος θυμίζει το επιθυμητό φωνήεν, δε φέρνει τόσο σε ανθρώπινη φωνή καθώς λείπει η “χροιά”. Το γεγονός αυτό, είναι πιθανό να οφείλεται στο ότι η κρουστική απόκριση συνελίσσεται με συρμό ώσεων δ[n] και όχι με συναρτήσεις sinc. Αυτό έχει ως αποτέλεσμα ο ήχος που παράγεται να είναι τελείως μονοκόμματος.

-Σχετικά με το liftering στον τομέα του Cerstrum ήταν καθοριστικής σημασίας το μήκος του παραθύρου, καθώς ακόμα και με μικρές αλλαγές σε αυτό, άλλαζε τελείως ο τελικός ήχος.