

A new frame-based statistical strategy for bird migration clutter removal in wind profiler radar data

Gerd Teschke

November 23, 2005

Abstract

This manuscript reports on the development of a new clutter removal routine for wind profiler radar data. The basic idea is to put together already existing building blocks from very different areas (such as redundant Gabor frame expansions and statistical test theory) that yield in its combination a very well suited tool for detecting and removing clutter components caused by strong bird migration. From the inverse problem perspective, we treat the problem of simultaneous signal decomposition and restoration. We finally present a completely working software prototype and a sequence of real life examples.

1 Introduction

In this manuscript we develop a new filter algorithm that is based on frame-based signal expansions and statistical test theory. The two essential ingredients come from [1] and [2]. The advantage in contrast to known filter techniques relies on the fact that a redundant time-frequency representation of a signal under consideration contains much more detailed information than a simple Fourier power spectrum (or a little sequence of them). The usefulness becomes clear when observing that in particular bird returns are very transient signal. On the other hand, because of the redundancy of the time-frequency representation, any constructed statistical test is much more confident, e.g. in [1] they use at most 36 observations to get proper estimates for specific test values. In our proposed situation we may use "as much as we want" - depending on the chosen time resolution of our phase space.

The organization of the manuscript is as follows: in Section 2.1 we collect the ingredients for our frame-based representation of the radar data, in Section 2.2 we explain the statistical filtering procedure, in Section 3.5 we explain the filtering method and sketch the data flow when applying the filtering routine to radar data, and finally in Section 3.1 we present a sequence of examples.

2 Mathematical Considerations

In what follows we briefly sketch the basic ideas behind our filter procedure.

2.1 Representation of radar data in phase space

In this section we focus on redundant discrete Gabor expansions. We shall mainly follow the work of J. Wexler and S. Raz, see [2], in which they have considered discrete phase space representations. In particular, the authors suggest a way of effectively computing such a redundant presentation and, by solving a biorthogonality relation, they provide a direct reconstruction algorithm.

For readers convenience, we briefly review the basic steps on how to arrive at a discrete Gabor representation of a given signal.

Let $x(t)$ be a given continuous time signal. A Gabor expansion is then given by

$$x(t) = \sum_{m,k} a_{m,k} h_{m,k}(t), \quad (1)$$

where

$$h_{m,k}(t) = h(t - mT) \exp ik\Omega t. \quad (2)$$

The computation of the frame coefficients relies on the so-called dual (or here the biorthogonal) window function γ ,

$$a_{m,k} = \langle x, \gamma_{m,k} \rangle = \int x(t) \bar{\gamma}_{m,k}(t) dt, \quad \gamma_{m,k}(t) = \gamma(t - mT) \exp ik\Omega t. \quad (3)$$

The parameter T controls the discrete linear shift along the time axis and Ω the sampling shift in the frequency domain. The choice of h , T and Ω directly controls the existence, uniqueness, convergence properties and the numerical stability of the expansion. A Relaxation of the classical constraint $\Omega T = 2\pi$ generates a crucial degree of freedom in the Gabor expansion. This at the expense of oversampling and a possible non-uniqueness for $\Omega T < 2\pi$ and a loss of stability for $\Omega T > 2\pi$.

Let us now switch to the discrete case which requires a few changes in (1). Assume we are given some discrete and periodic time signal $x(n)$, $n = 0, \dots, N-1$, and, moreover, $\Omega T < 2\pi$. Then, let M, K, \bar{M}, \bar{K} and N be such that

$$N = M\bar{K} = \bar{M}K. \quad (4)$$

The critically sampled situation is in the subsequent analysis still covered when choosing $M = \bar{M}$ and $K = \bar{K}$. Since we deal with periodic signals, we have to periodize the analysis and synthesis windows,

$$\tilde{h}(n) = \sum_l h(n + lN) \quad , \quad \tilde{\gamma}(n) = \sum_l \gamma(n + lN).$$

Under these assumptions $x(n)$ can be represented by

$$x(n) = \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} a_{m,k} \hat{h}_{m,k}(n), \quad (5)$$

where the Gabor coefficients can be derived from

$$a_{m,k} = \sum_{n=0}^{N-1} x(n) \tilde{\gamma}_{m,k}(n) \quad (6)$$

and where with $W = \exp(2\pi i/K)$

$$\hat{h}_{m,k}(n) = \tilde{h}(n - m\bar{M})W^{nk}, \quad \hat{\gamma}_{m,k}(n) = \tilde{\gamma}(n - m\bar{M})W^{nk}.$$

In order to ensure perfect reconstruction, the sequences \tilde{h} and $\tilde{\gamma}$ have to fulfill the following biorthogonality relation

$$\sum_{n=0}^{N-1} [\tilde{h}(n + m\bar{M}) \exp(2\pi i \bar{M}nk)] \tilde{\gamma}(n) = N/MK \delta_m \delta_n$$

$$0 \leq m \leq \bar{M} - 1, \quad 0 \leq k \leq \bar{K} - 1, \quad (7)$$

or, in matrix notation,

$$H\tilde{\gamma} = v, \quad (8)$$

where $v^t = (N/MK, 0, \dots, 0)$ is a vector of length $\bar{M}\bar{K}$, $\tilde{\gamma}^t = (\tilde{\gamma}(0), \dots, \tilde{\gamma}(N-1))$ and H is a matrix of dimension $\bar{M}\bar{K} \times N$ which is organized as a $\bar{M} \times \bar{M}$ block matrix where the size of each individual block is $\bar{K} \times \bar{K}$.

For our case of interest, i.e. oversampling situation, system (8) is underdetermined, leading to a nonunique $\tilde{\gamma}$, and consequently, to a nonunique Gabor representation of $x(n)$. The nonuniqueness can be used to add further desirable constraints on the solution, e.g. minimum energy. In this particular case the solution is often called the canonical dual window function and can be derived by

$$\tilde{\gamma} = H^t(HH^t)^{-1}v.$$

Once we have computed for some given \tilde{h} some adequate $\tilde{\gamma}$, we may compute by (6) the Gabor coefficients of $x(n)$ and we may reconstruct $x(n)$ thanks to (5).

Thus, the general data flow for a filtering (clutter removal) routine in Gabor phase space can be performed as illustrated in Figure 1. It remains to find/develop an adequate filtering technology.

2.2 Statistical filtering

In order develop a filtering mechanism, we have to search for a model or something adequate that describes the component to be extracted very well. Our particular goal is to identify echoes coming from fliers. The resulting process of this echo component is rather complicated (some transient process), but signals due to weather and to the radar system can be assumed to be Gaussian (where we note that the former has a colored Fourier power spectrum (short range dependency) and latter a white Fourier power spectrum). I.e. when we are able to make use of these descriptions and to find an good estimate for weather and the noise component, we

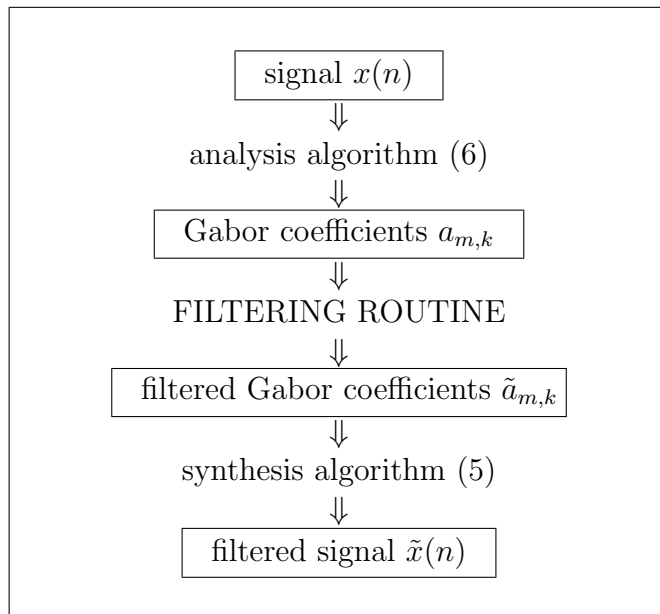


Figure 1: Data flow: from time signal representation to Gabor phase space representation and back (with a potential filtering step).

have some tool to isolate the fliers (such as birds) echoes. Here we will pick up some ideas presented in [1] and hope to find with them a well suited estimate for the "bird" signal component.

Let us assume that the statistics of the Gabor coefficients is nearly the same as for the ones of the corresponding Fourier power spectrum (this needs to be checked analytically, empirically this seems to be true but it is clear that the introduced redundancy may cause additional dependencies). Then we may argue as in [1] but apply the idea to the Gabor phase space representation: for each individual fixed index m we may consider the vector $|a_{m,k}|^2$. This vector represents a time localized Frequency spectrum, i.e. we have a total number of M such spectra. Now we may build order statistics $|a_{[m],k}|^2$, i.e. for all k ($k = 0, \dots, K - 1$) one has a descending order

$$|a_{[m]+1,k}|^2 \leq |a_{[m],k}|^2.$$

Then, an averaged spectral estimate (mean value) is

$$\hat{a}_k^{M_k} = \frac{1}{M_k} \sum_{m=0}^{M_k-1} |a_{[m],k}|^2, \quad k = 0, \dots, K - 1. \quad (9)$$

Note that the number of averaged spectra M_k can be chosen different for each k (but, for simplicity, here is not). The number M_k determines the subset of spectra under consideration. The Gaussian statistical test implemented in [1] is of the form

$$\frac{\text{var}(|a_{[m],k}|^2 \mid M_k)}{(\hat{a}_k^{M_k})^2} \leq 1, \quad (10)$$

and was used to discriminate between radar system noise and atmospheric radar return. Here the variance and mean computation uses the same subset of M_k . But actually, the $x = I + iQ$

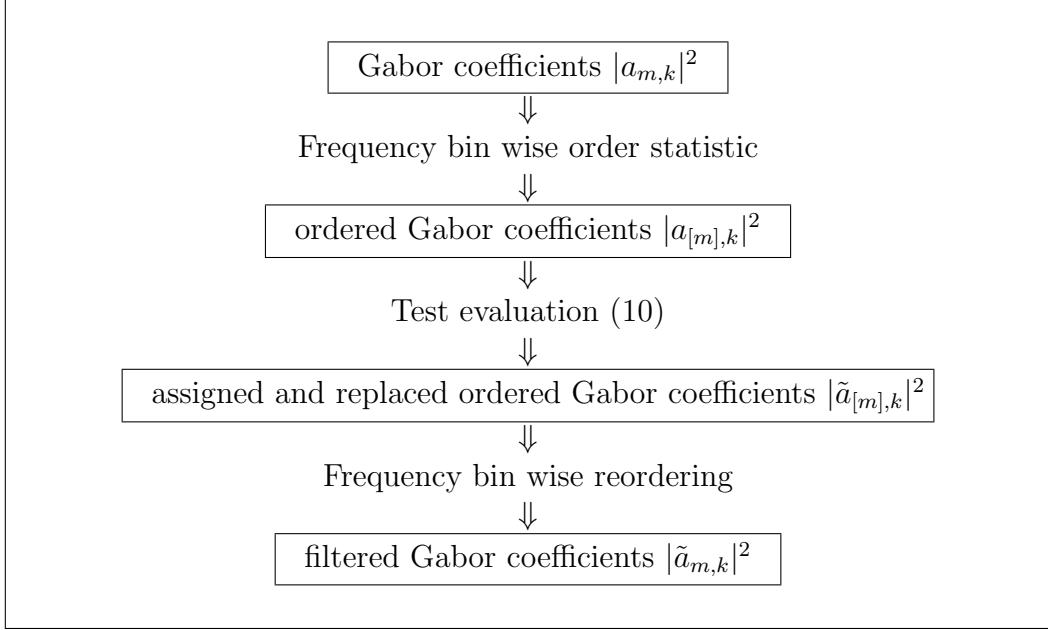


Figure 2: Data flow: statistical detection of "bird" - Gabor coefficients and its removal).

time series under consideration here are assumed to be Gaussian distributed for both noise and atmosphere. The power measured by the spectral estimates (here the Gabor coefficients) must be (nearly) exponentially distributed. Hence, the suggested test made in [1] was also a successful test for exponentially distributed data. This becomes clear since the test is incomplete because it relates only the first two moments and covers thus a wider class of distributions (but it was shown that this particular test was sufficient for the discrimination purpose). To apply this test to identify fliers, we recall that radar system noise and clear-air return from the atmosphere are in general much weaker than bird echoes. Thus, only the weakest signals in each spectra bin satisfying this test (10) are identified with noise or atmosphere. The stronger Gabor spectral estimates are then assumed to belong to fliers like birds.

Summarizing, to detect Gabor coefficients that belong to "bird" echoes we proceed as follows: starting from the Gabor phase space representation $|a_{m,k}|^2$ we build at first the order statistic for each individual frequency bin, $|a_{[m],k}|^2$ (in descending order for all $k = 0, \dots, K - 1$). Compute then test (10) for each individual k with decreasing number M_k of ordered Gabor coefficients and assign those coefficients that have values larger than 1 as coefficients belonging to "bird" echoes. Replace the assigned coefficients with a frequency bin wise mean value or zero (here we use the mean value replacement in order to preserve the atmospheric signal energy). The data flow is also illustrated in Figure 2.

3 Algorithmic Realization

The overall algorithmic realization is done by putting all ingredients together, see the overall data flow in Figure 3.

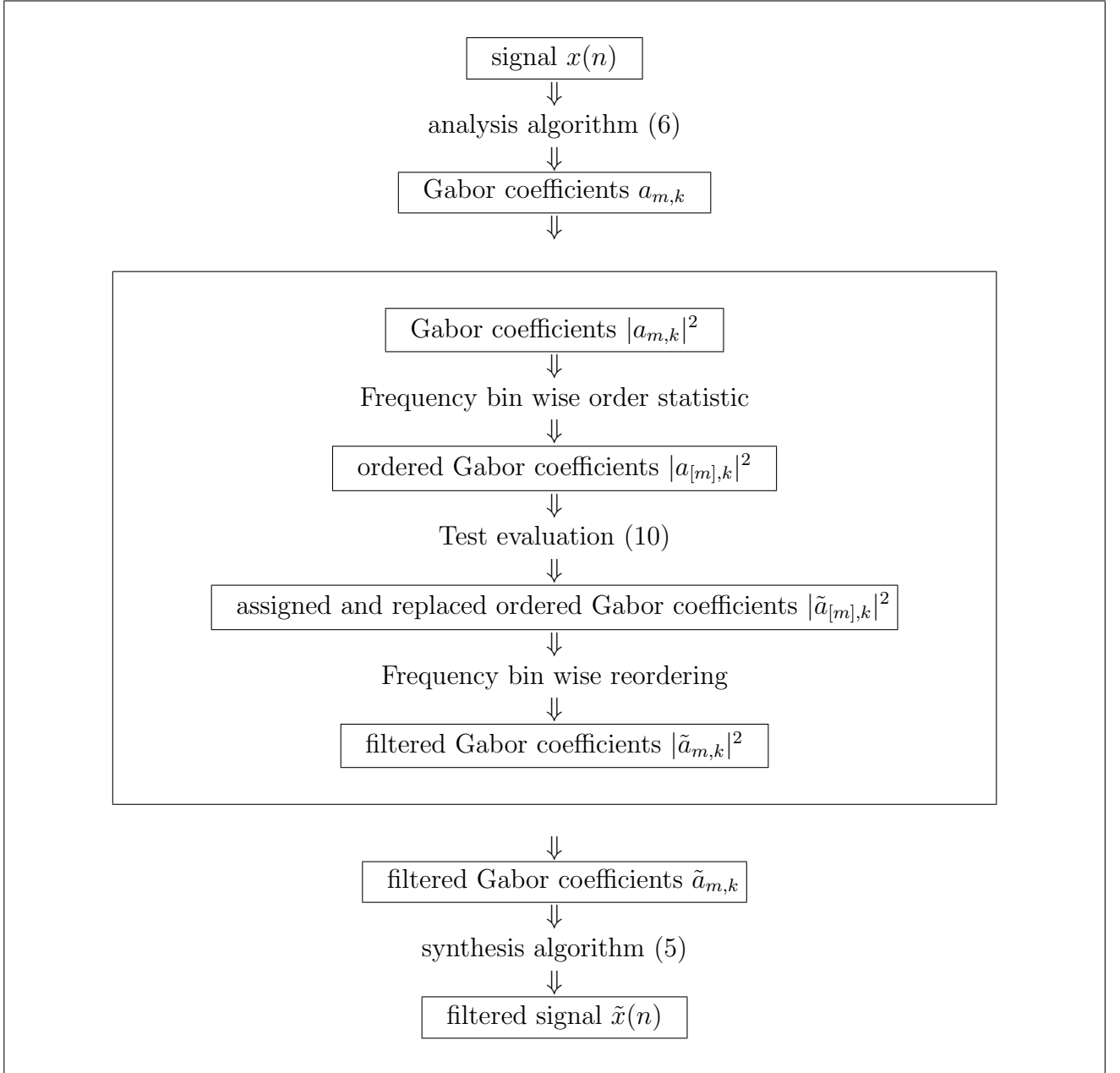
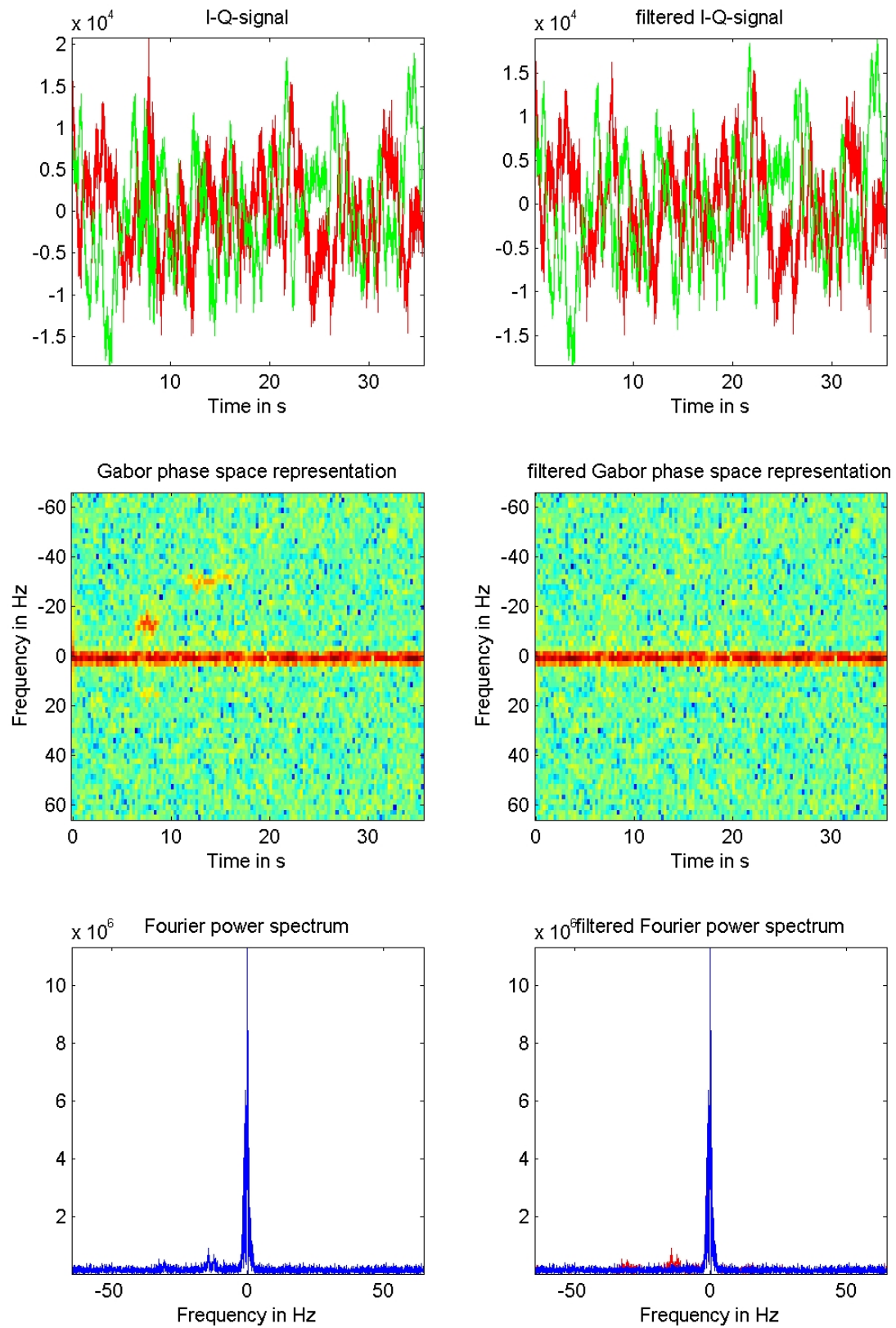


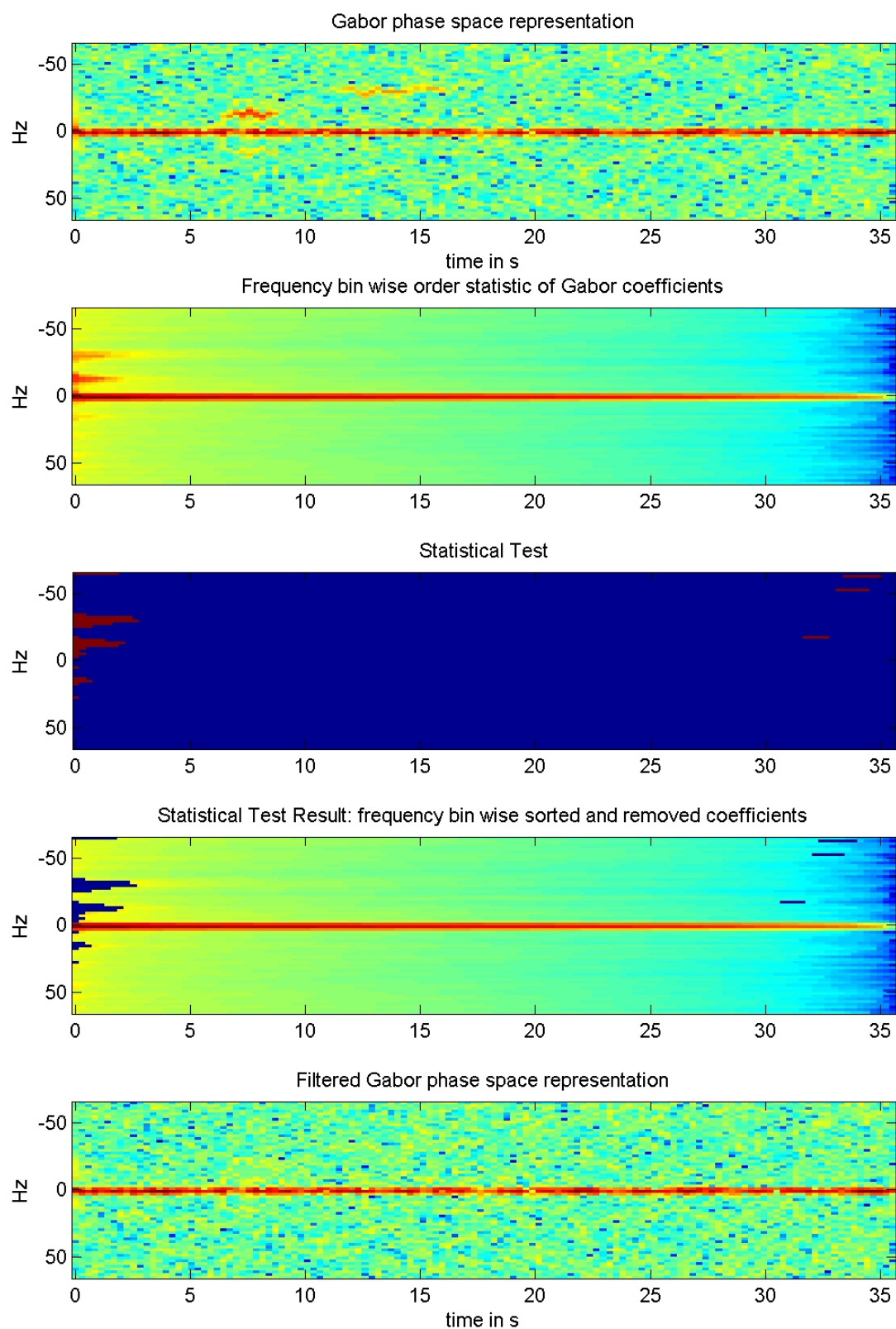
Figure 3: Data flow of the new TELE - filter algorithm.

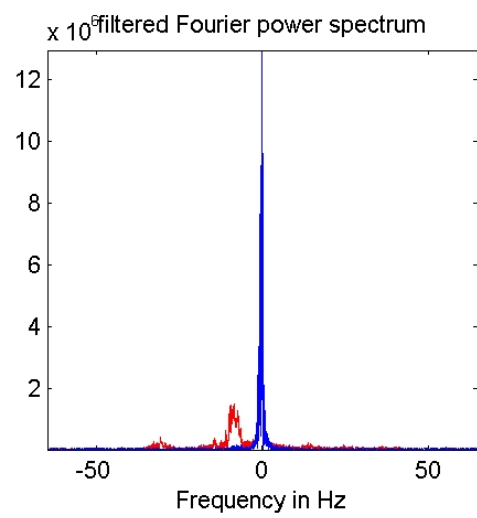
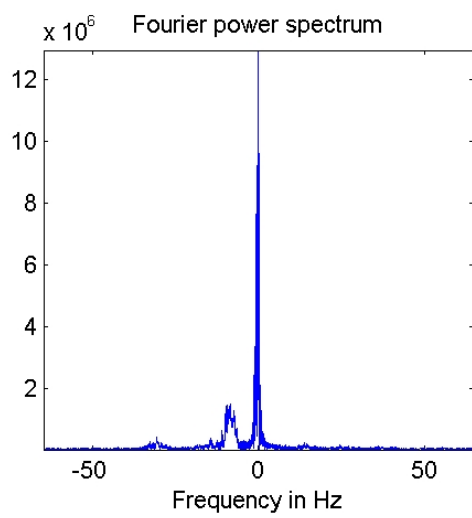
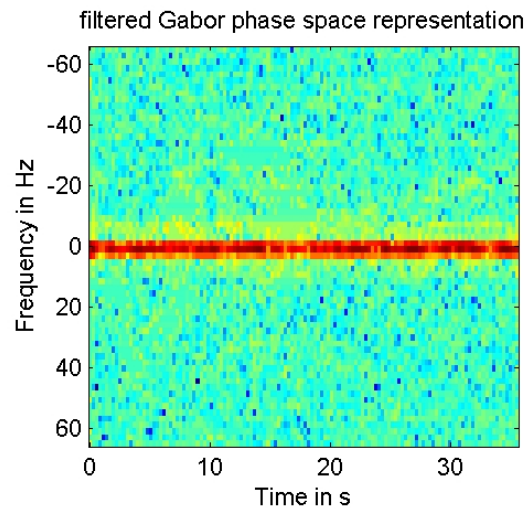
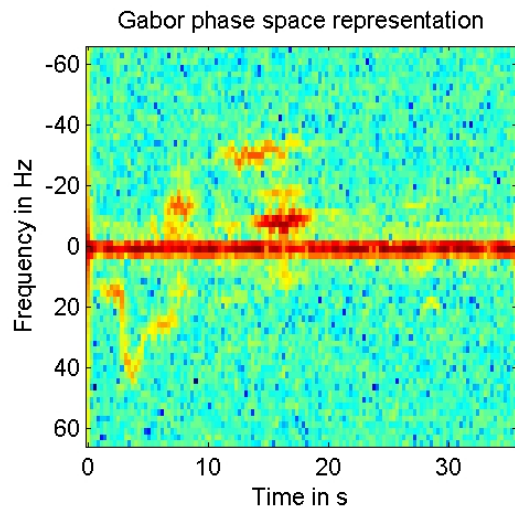
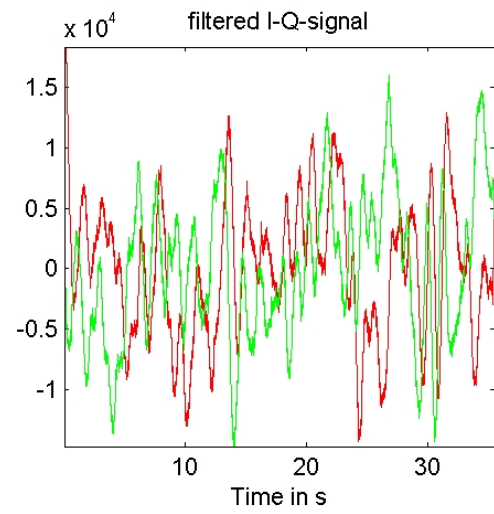
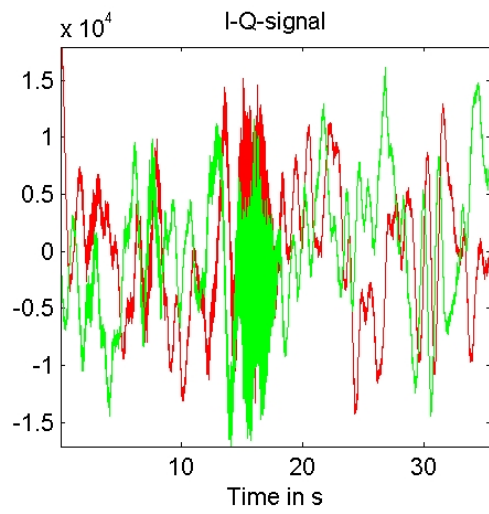
3.1 Convincing Examples

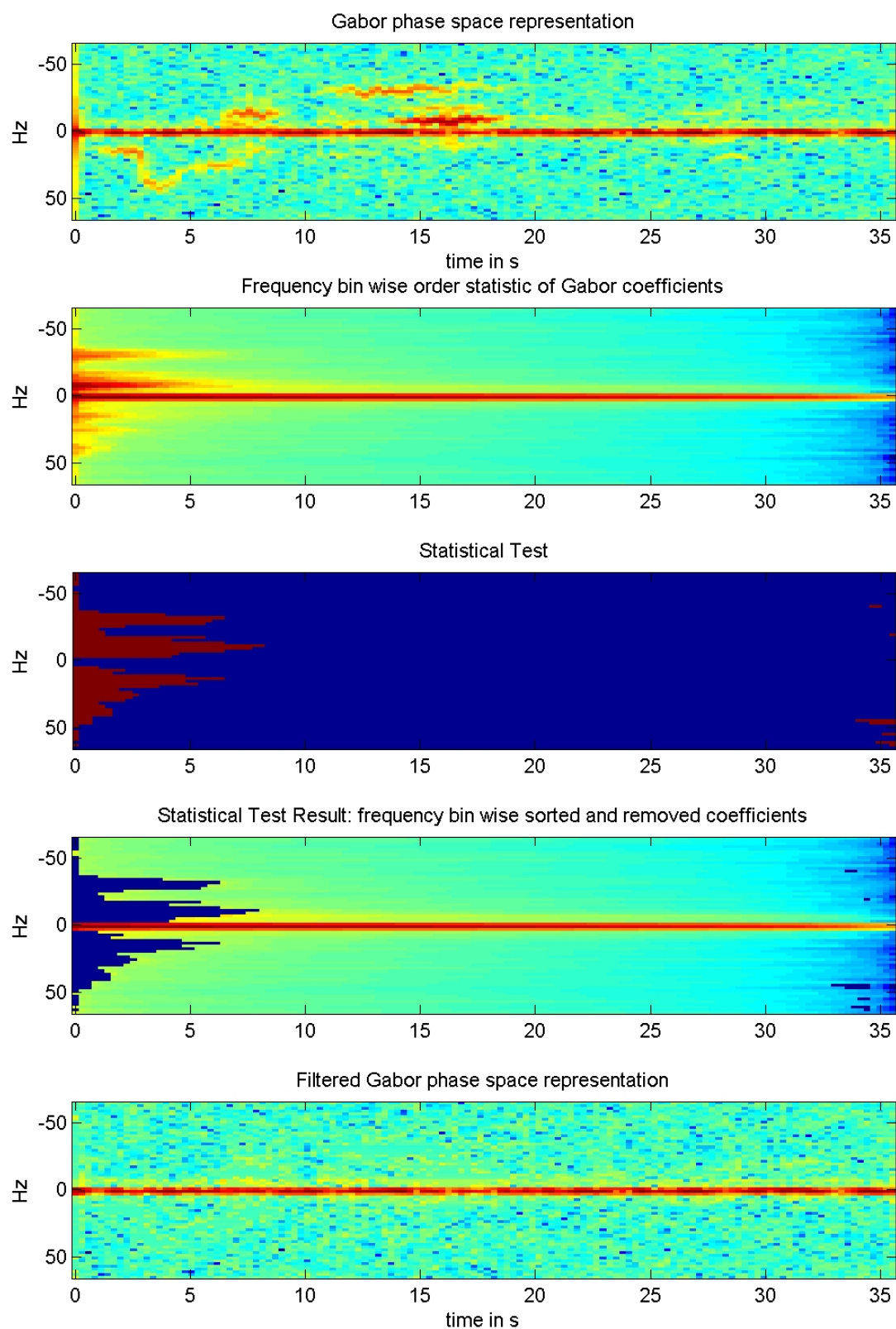
Data set: D05265

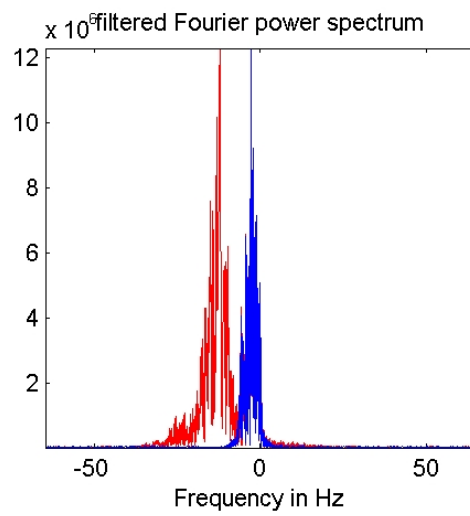
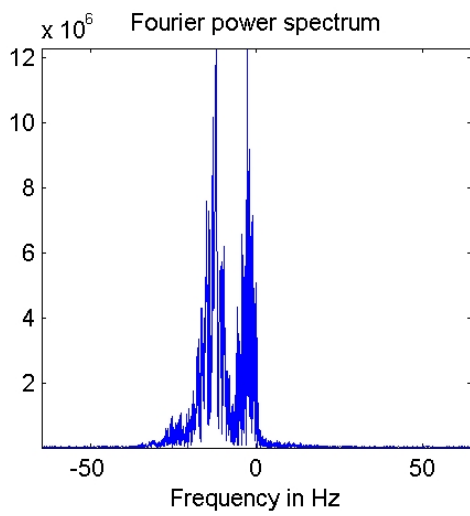
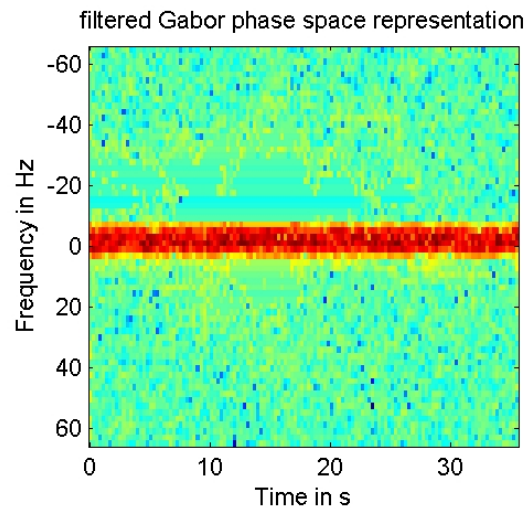
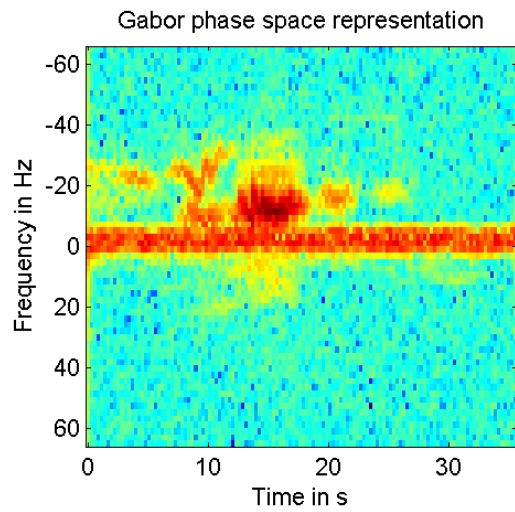
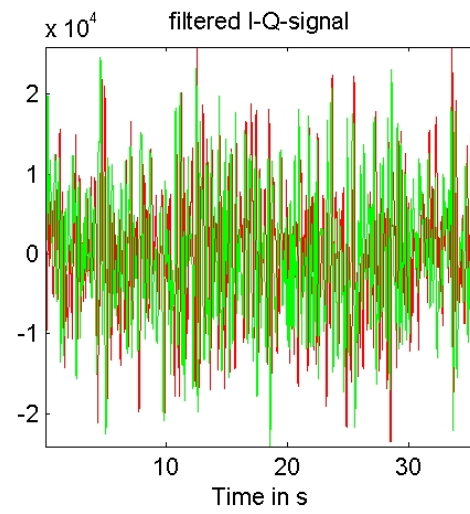
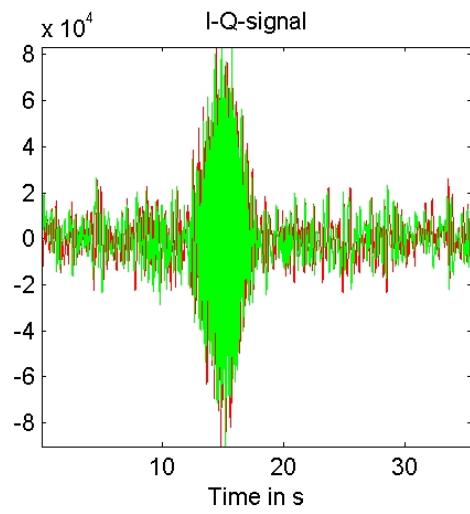
3.2 Dwell 1, Ranges 1–30, Detailed shown range 1, 2, 6, 14

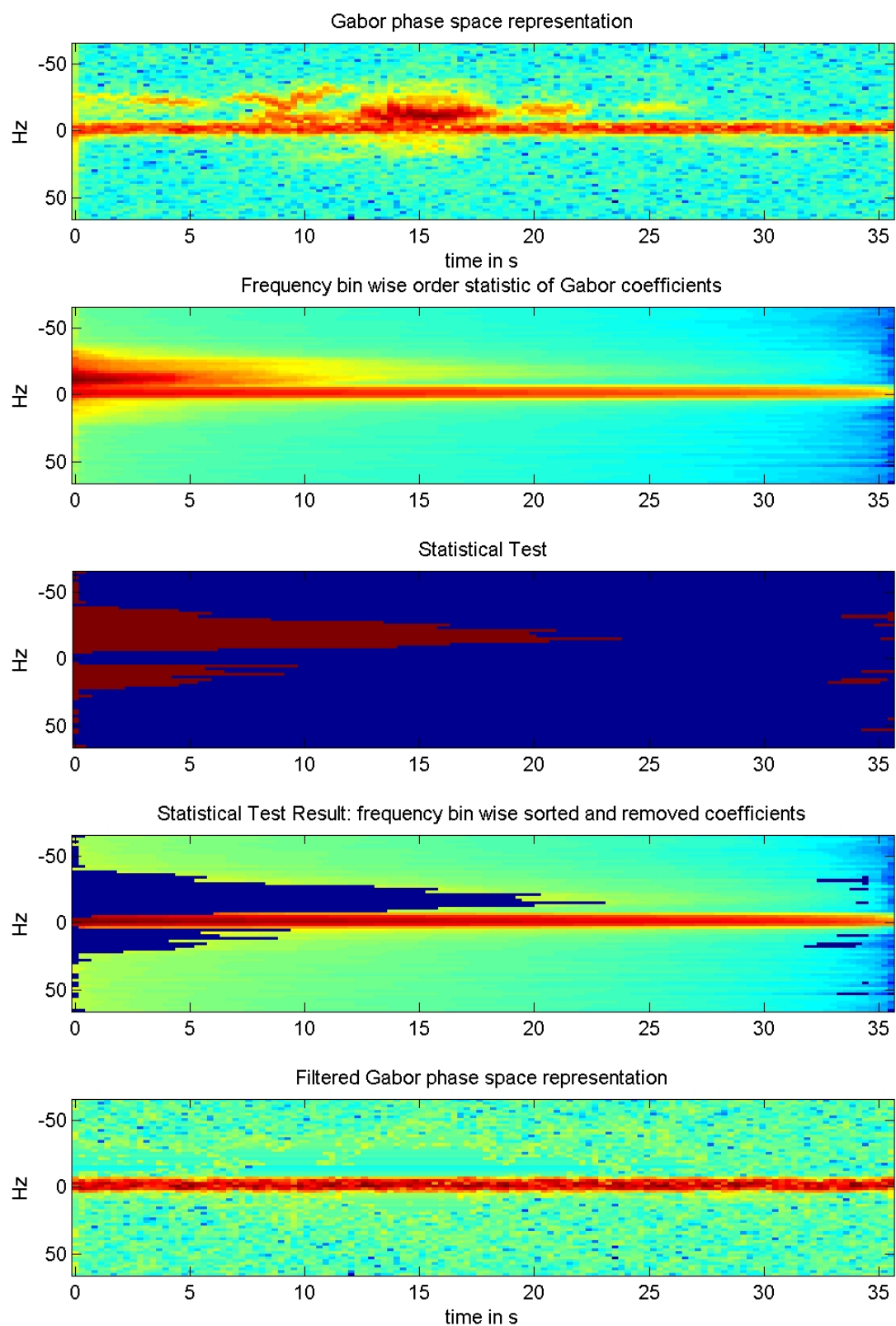


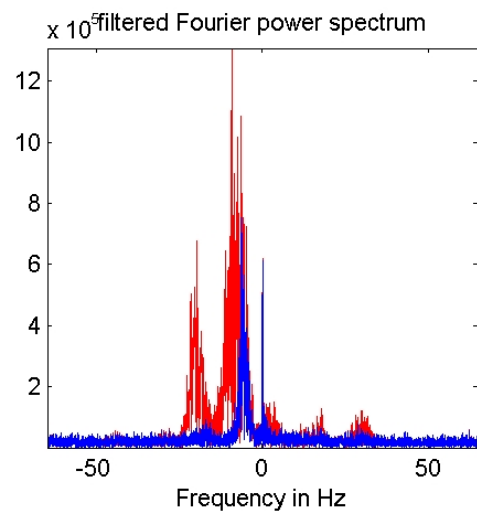
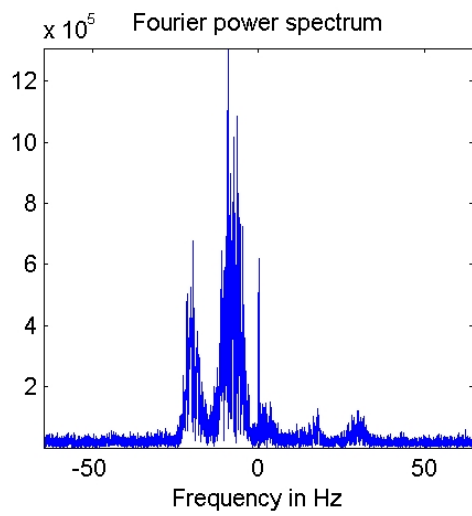
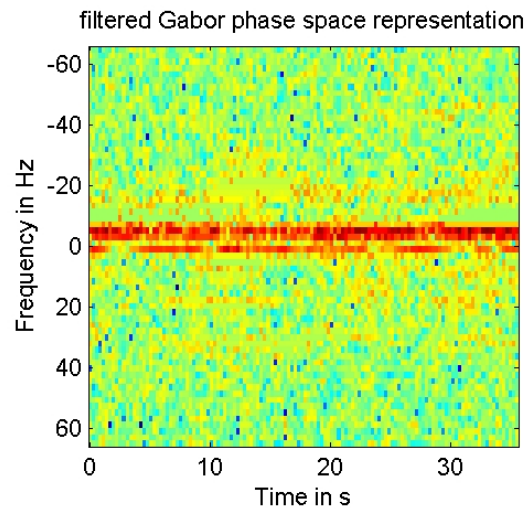
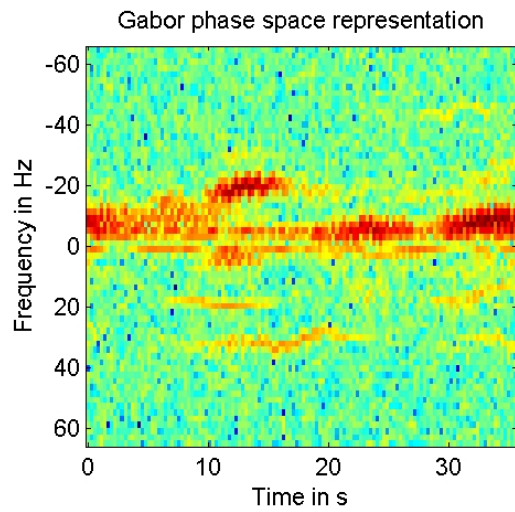
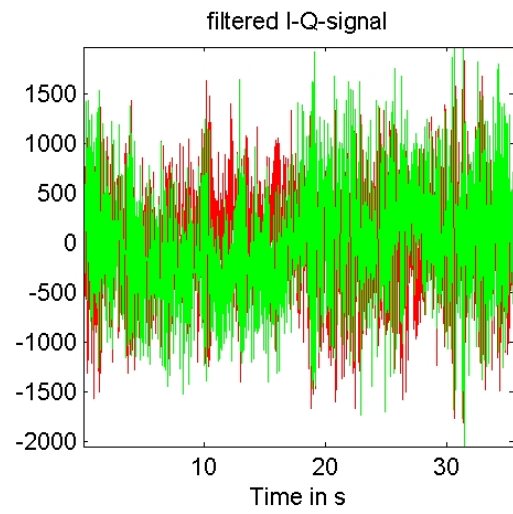
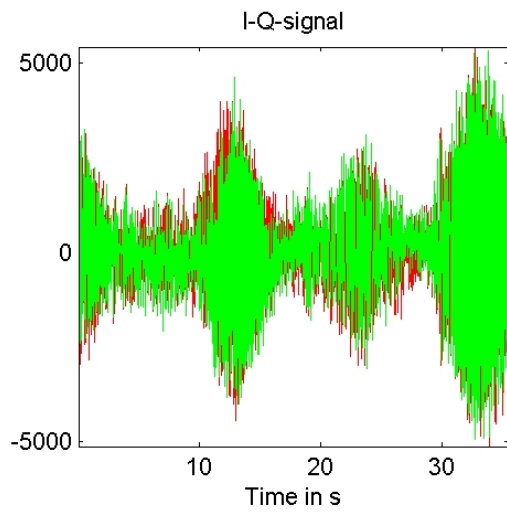


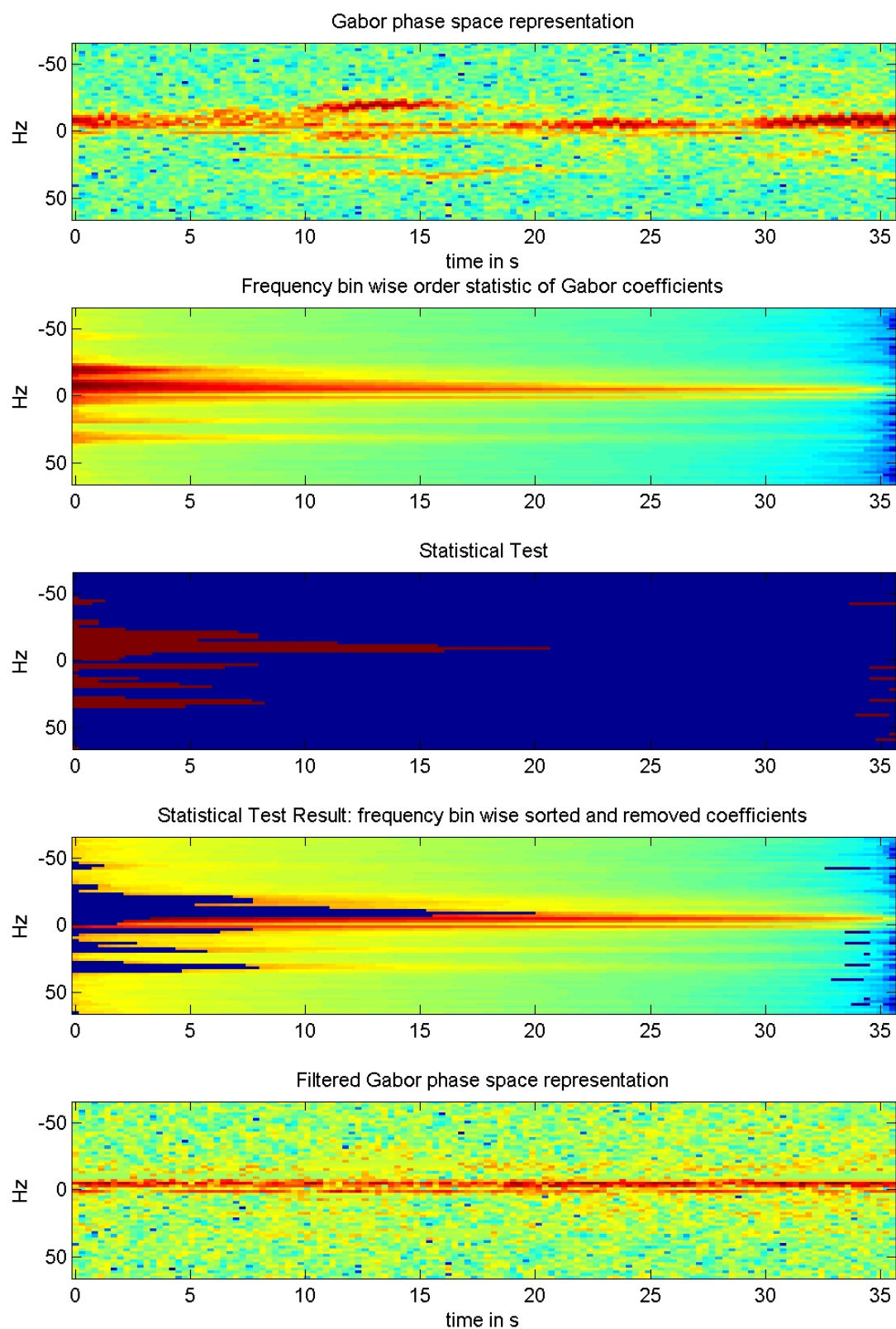


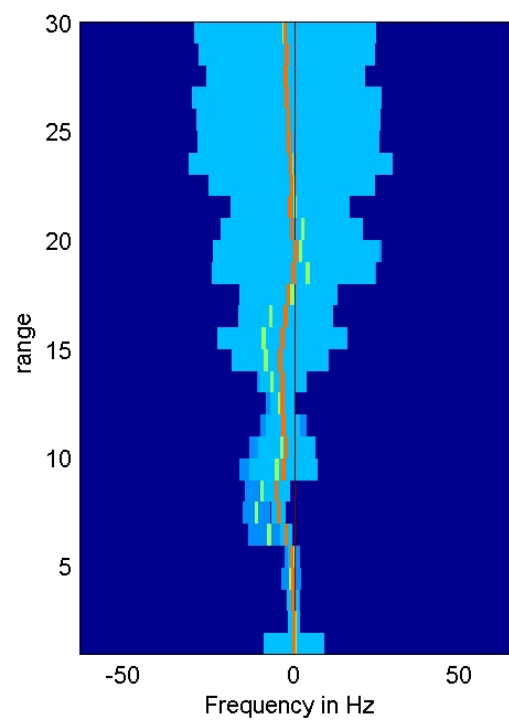
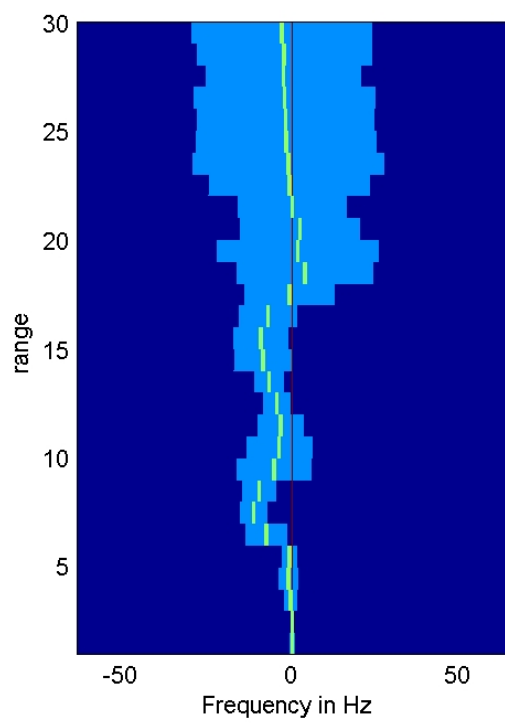
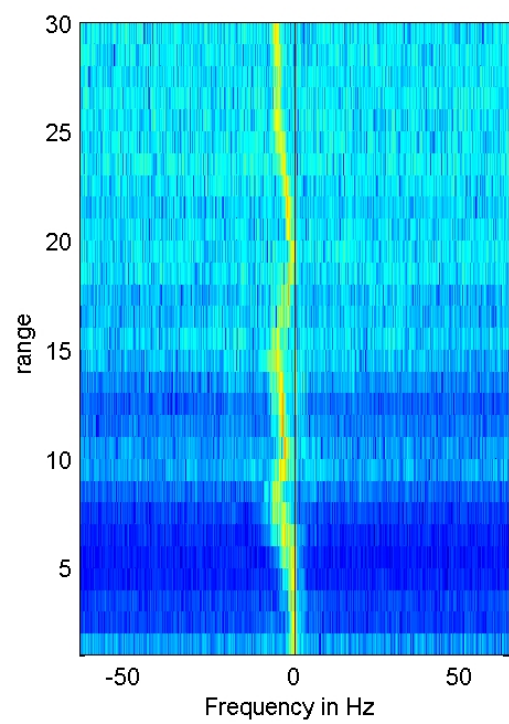
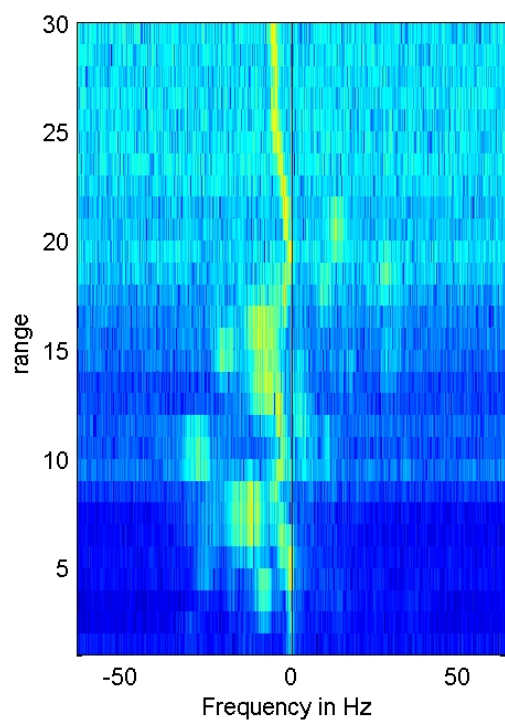




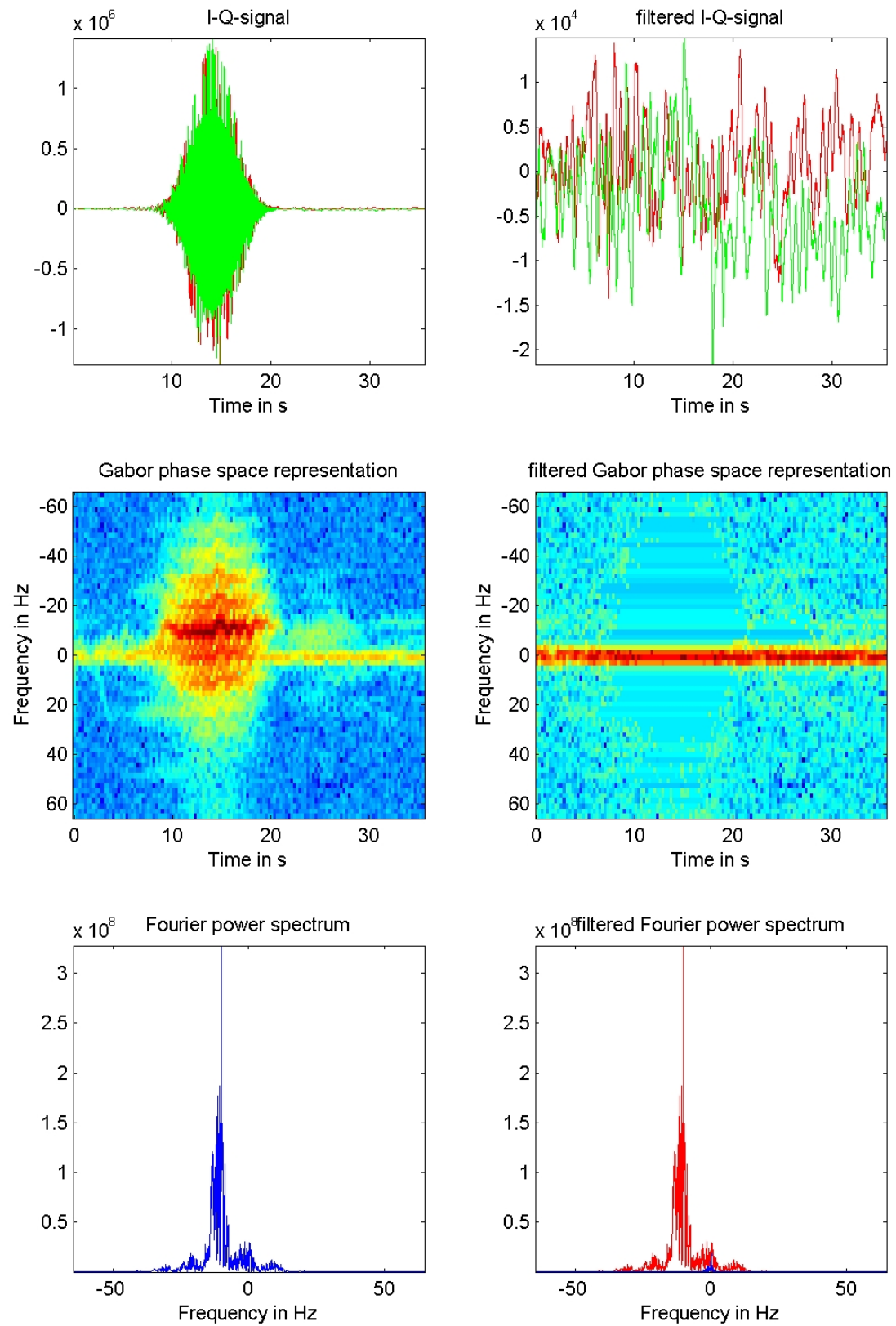


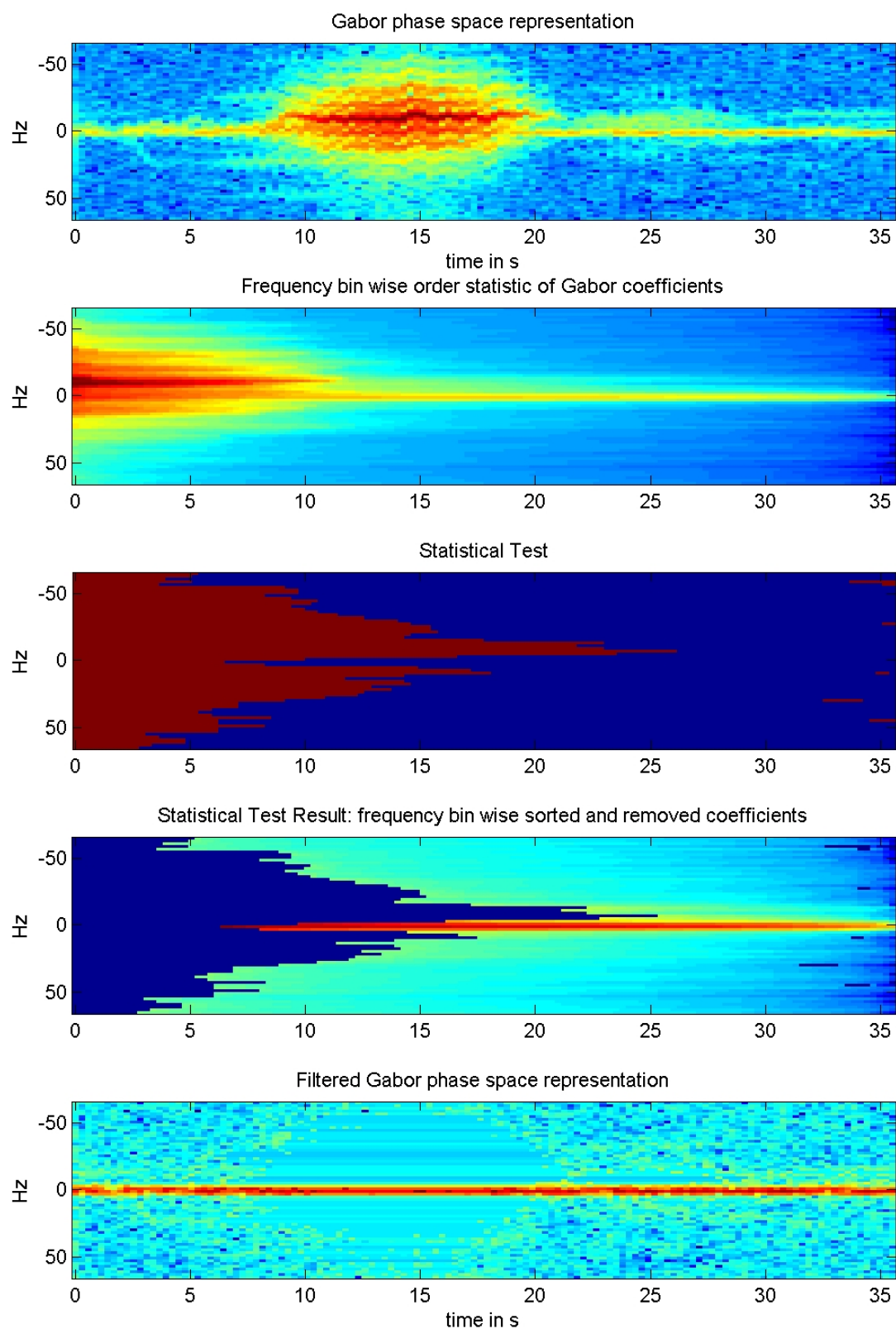


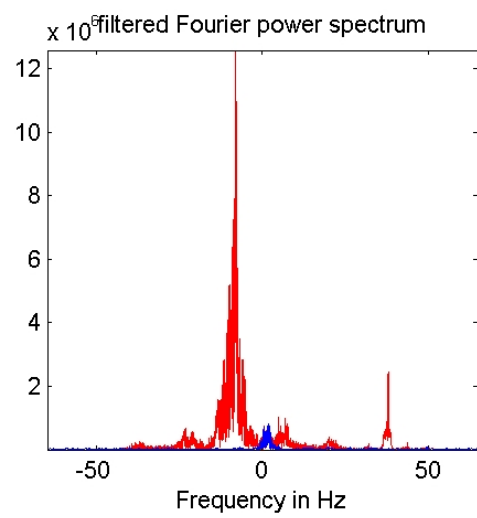
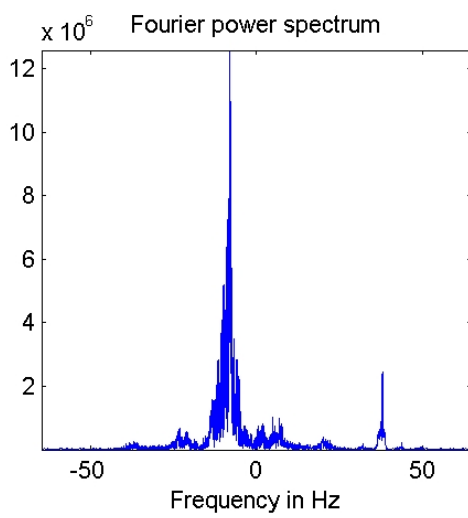
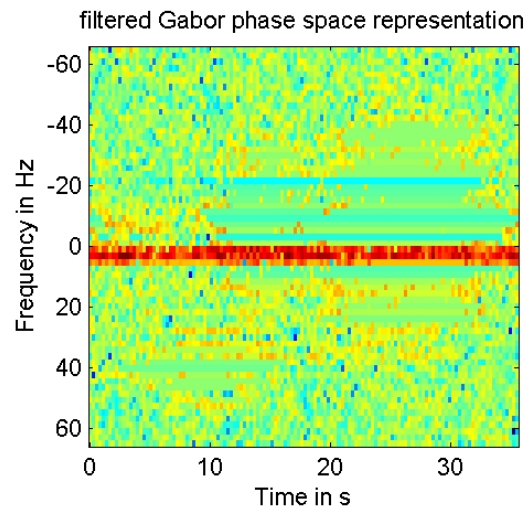
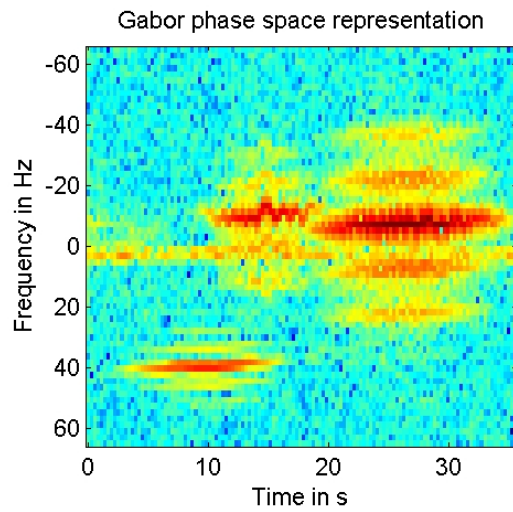
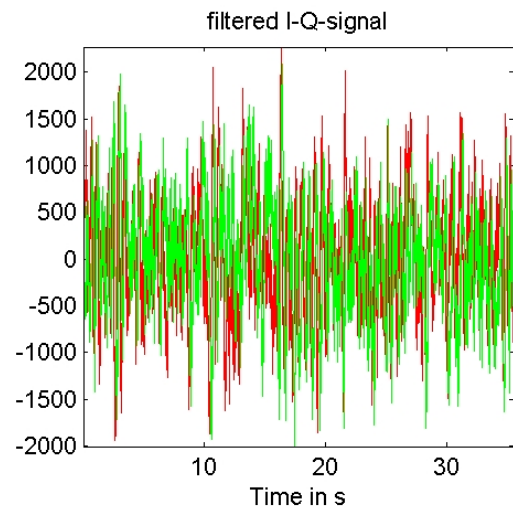
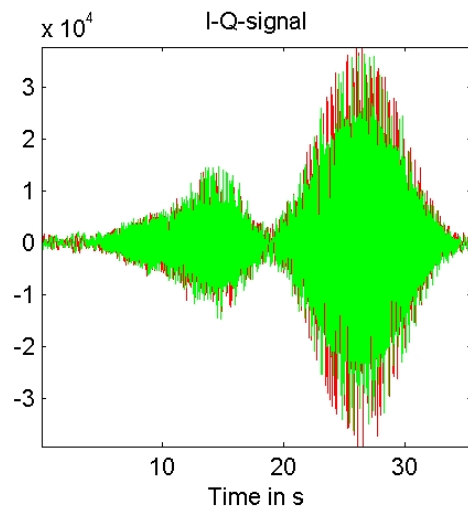


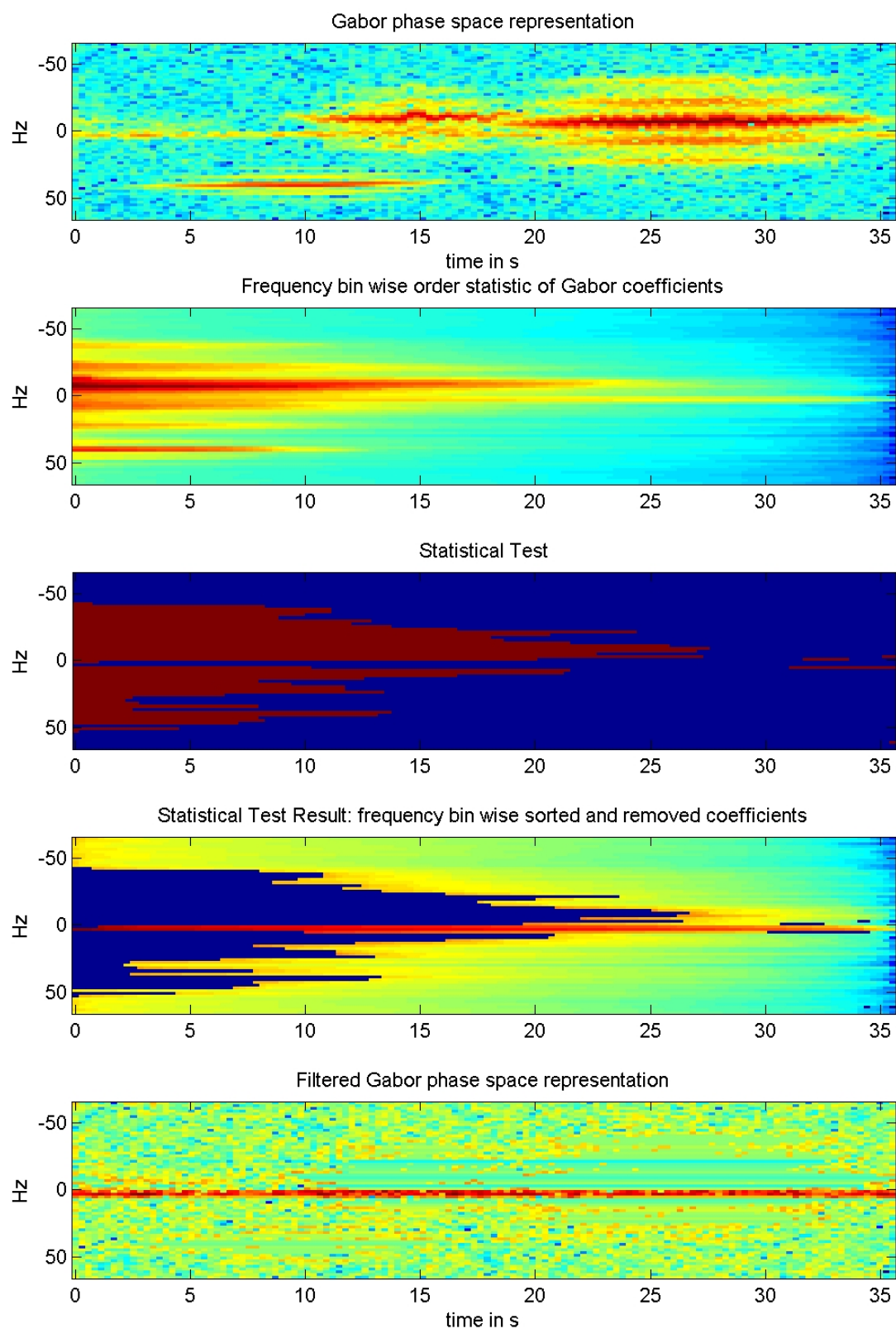


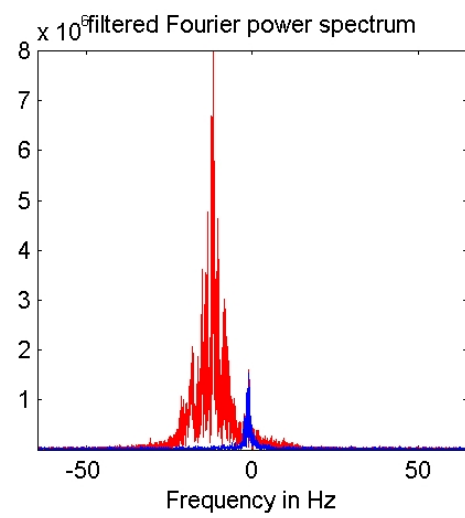
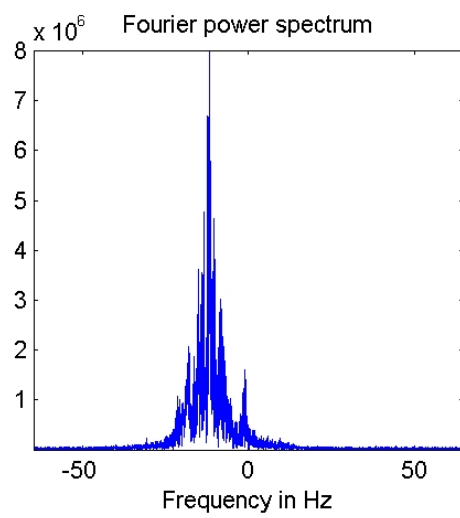
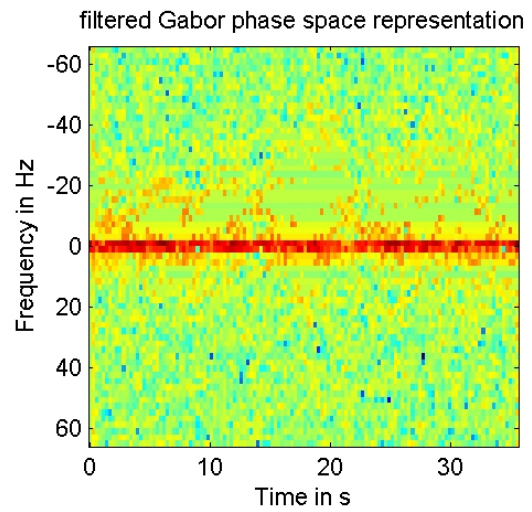
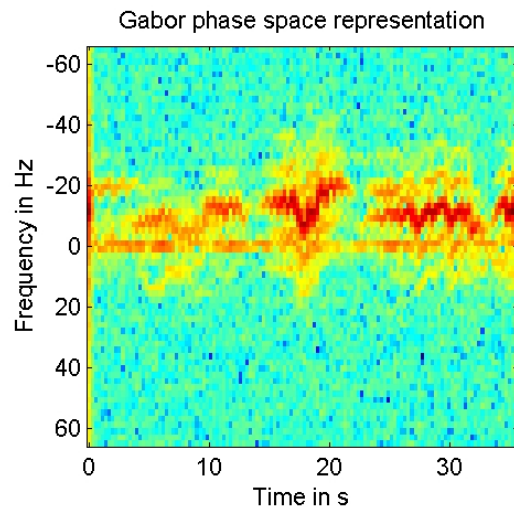
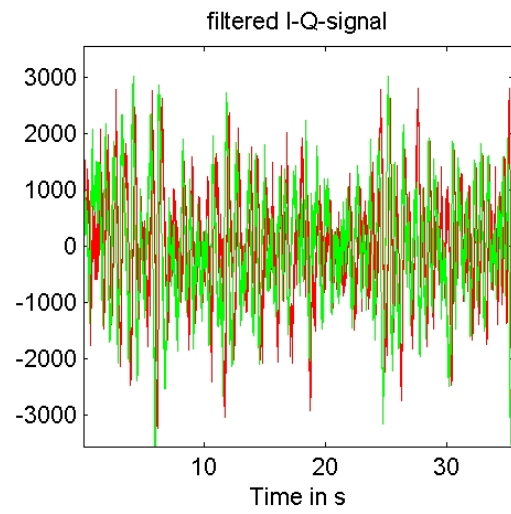
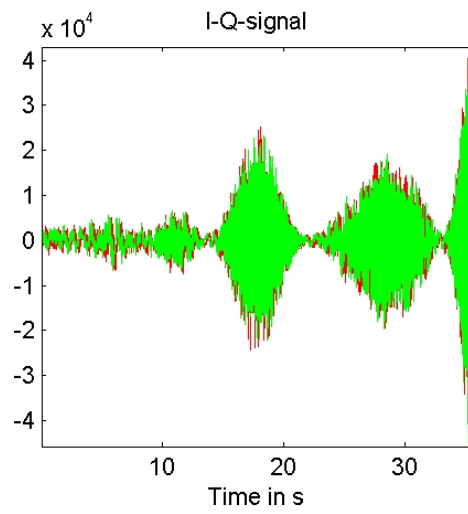
3.3 Dwell 10, Ranges 1–30, Detailed shown range 7, 9, 13, 19

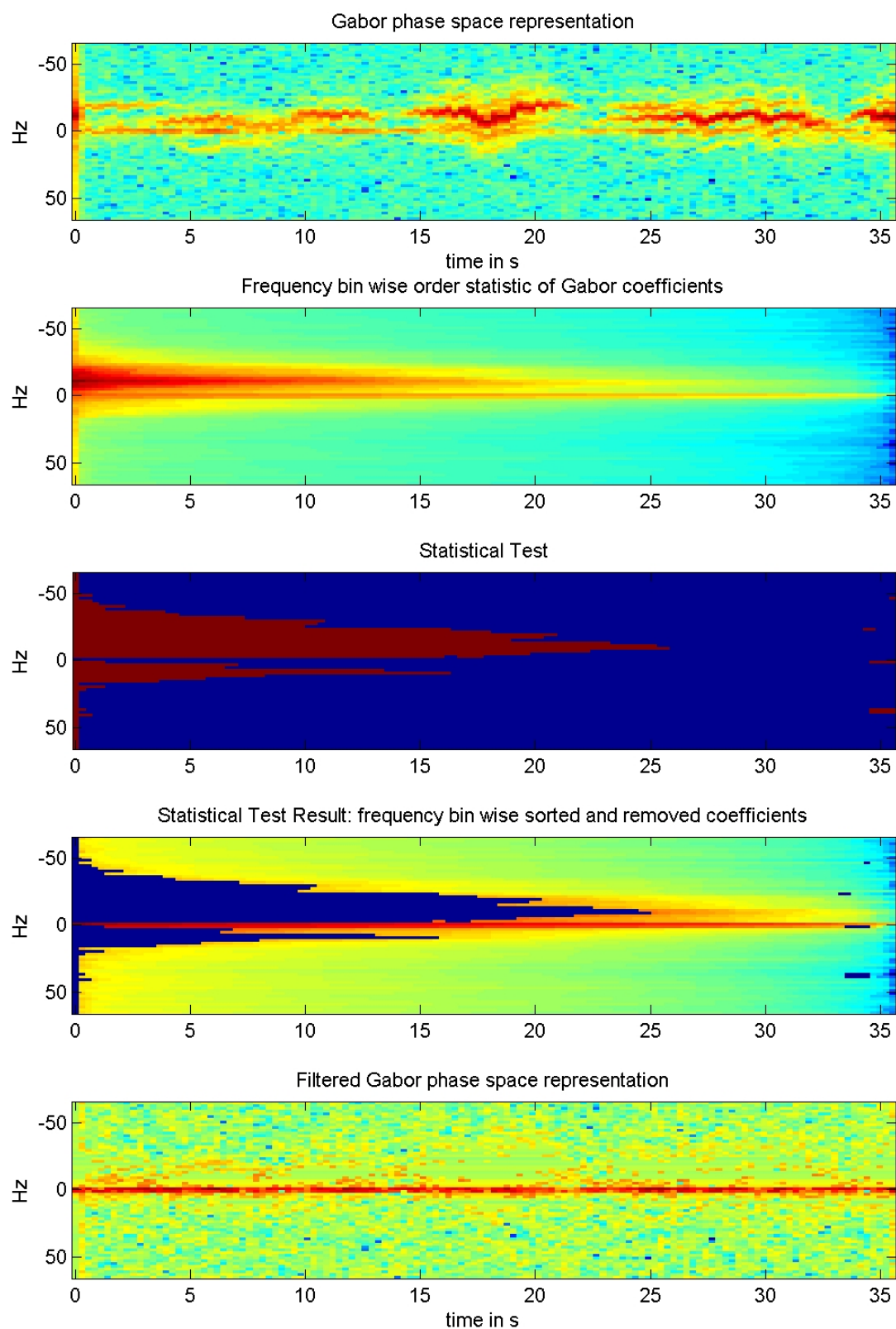


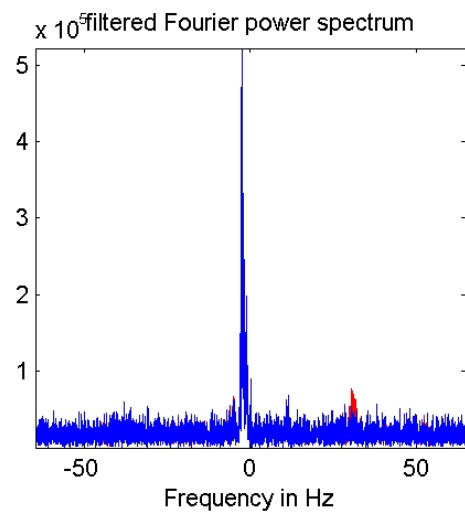
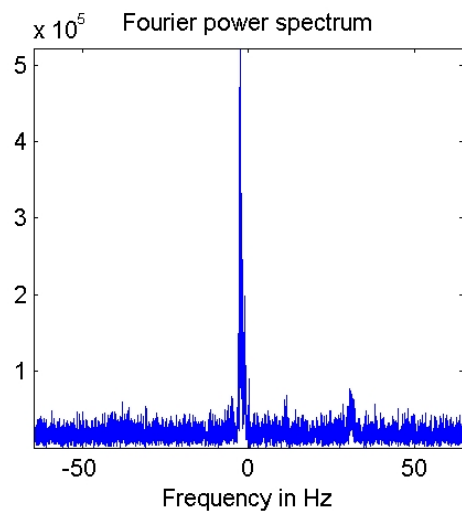
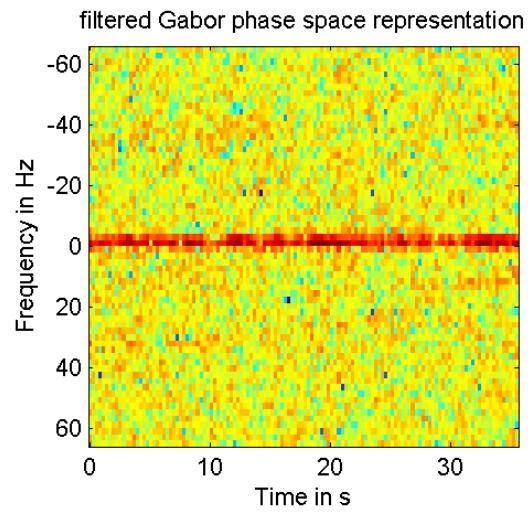
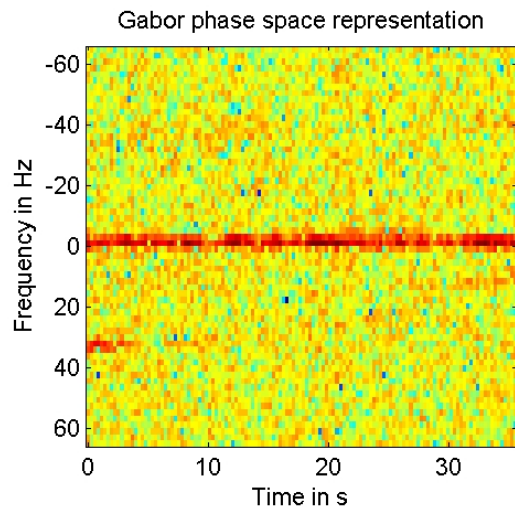
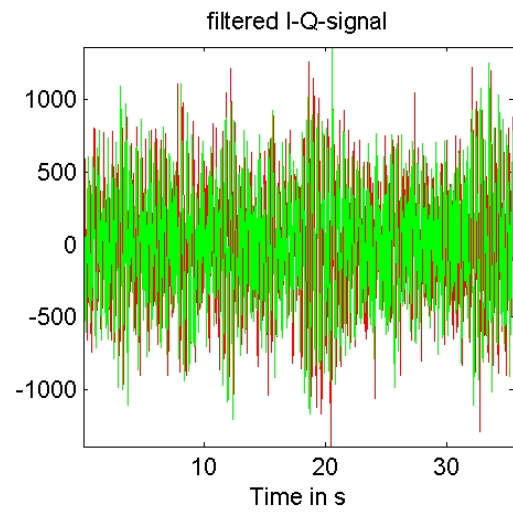
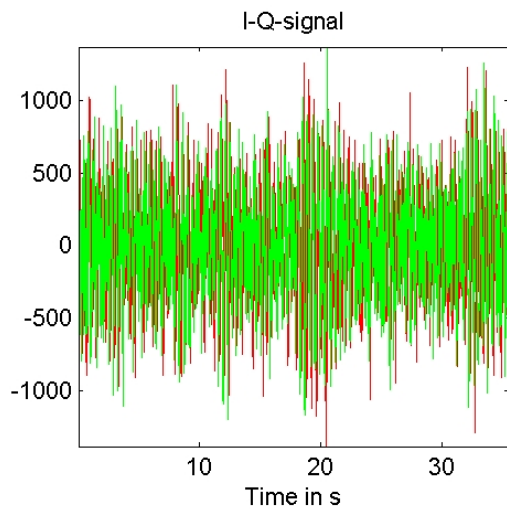


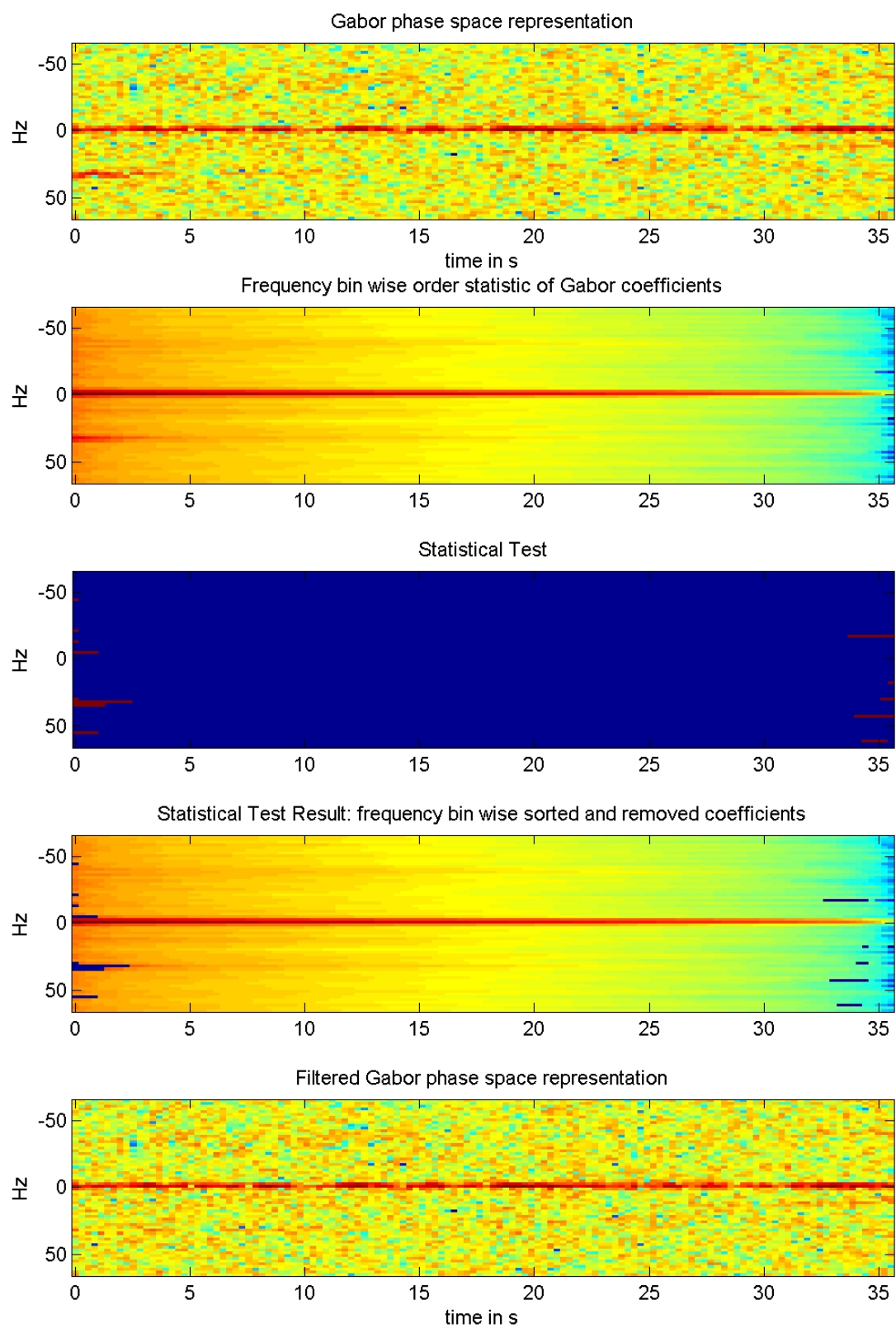


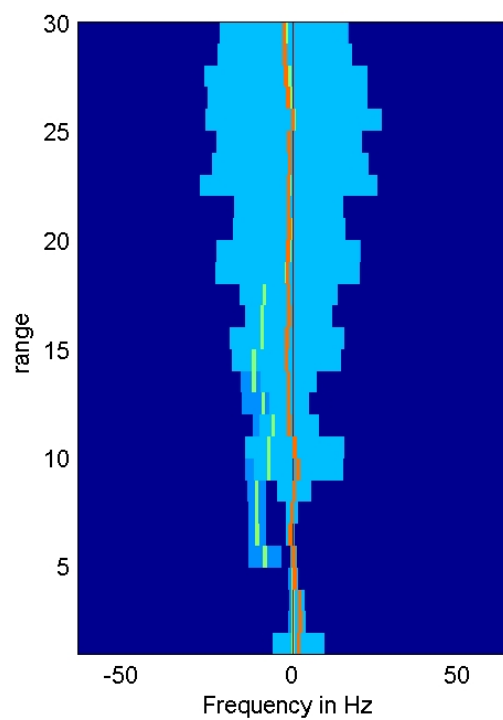
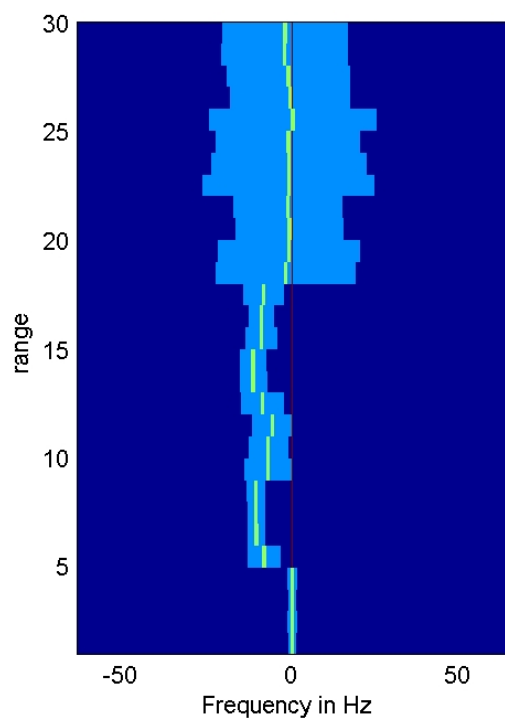
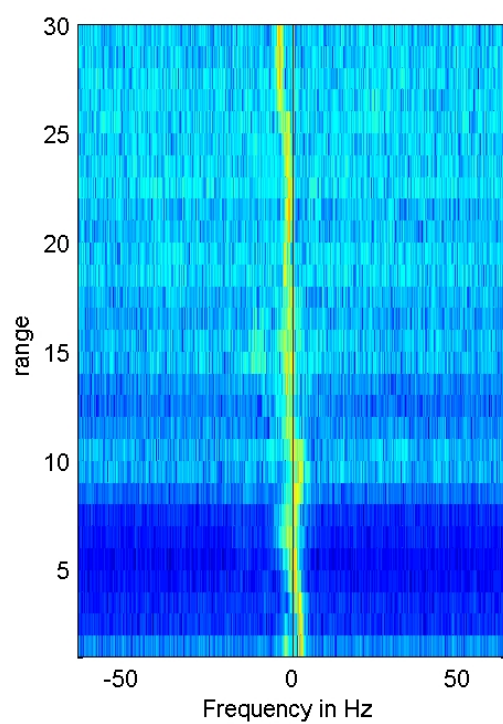
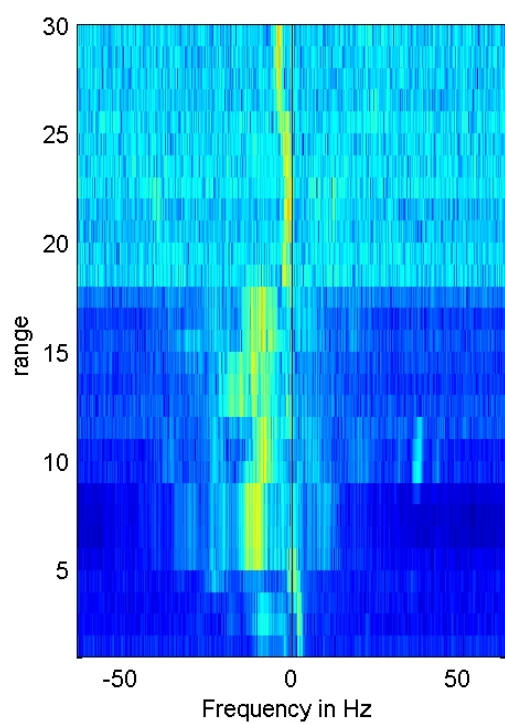




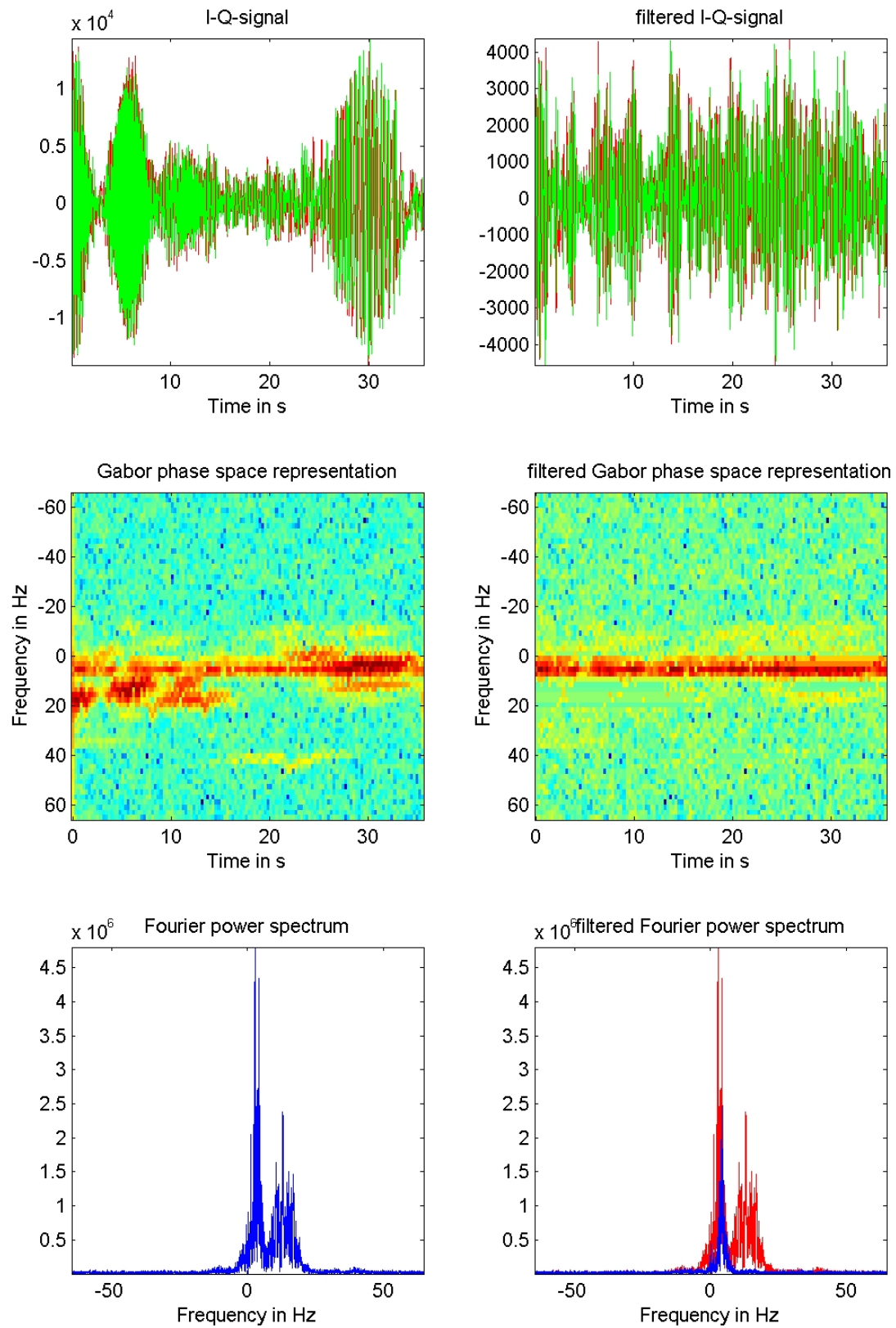


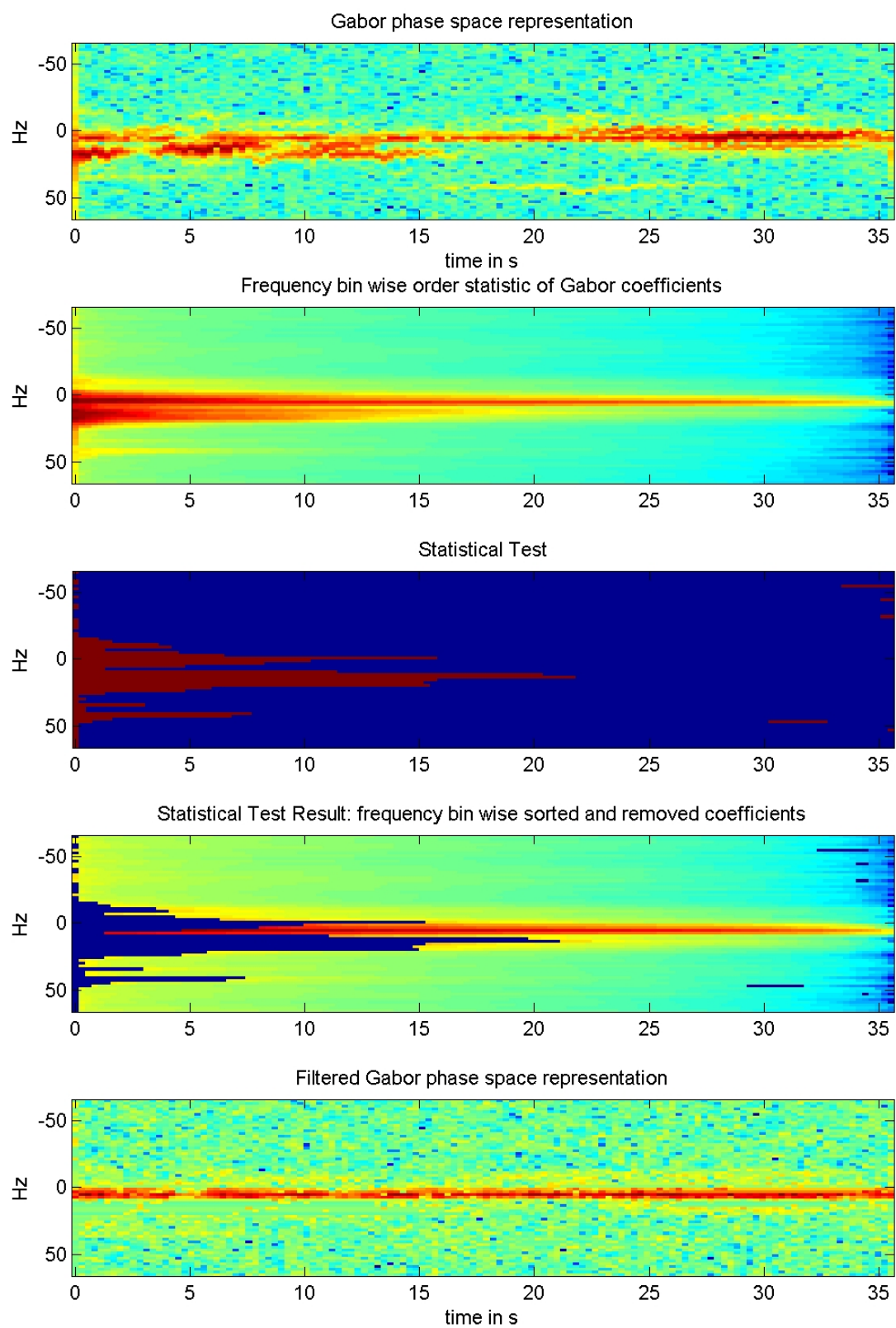


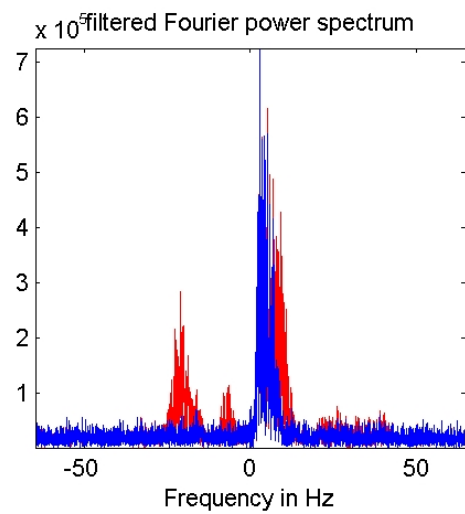
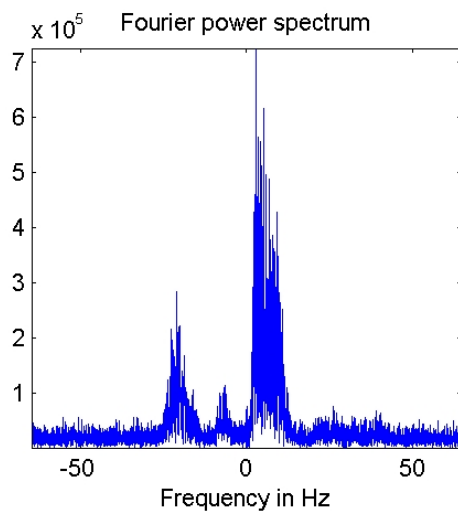
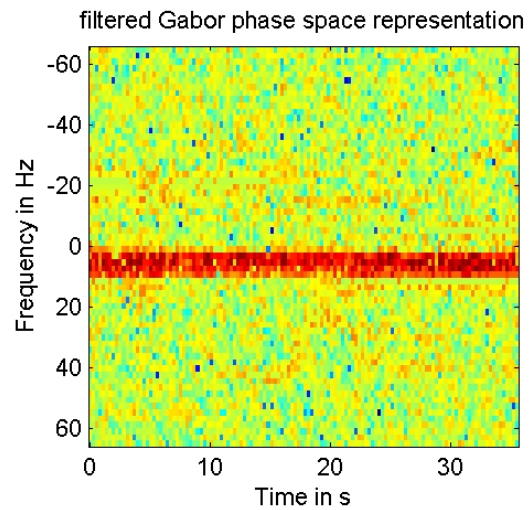
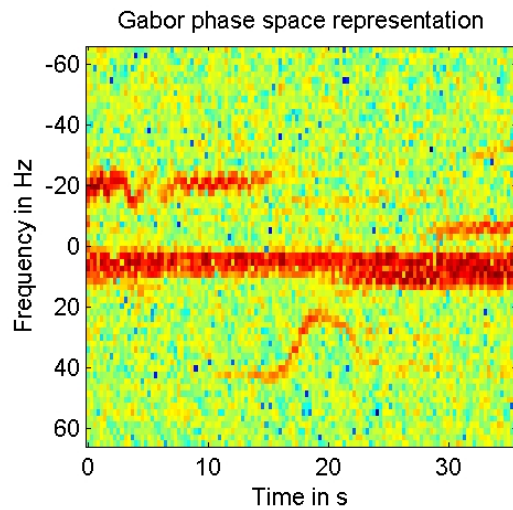
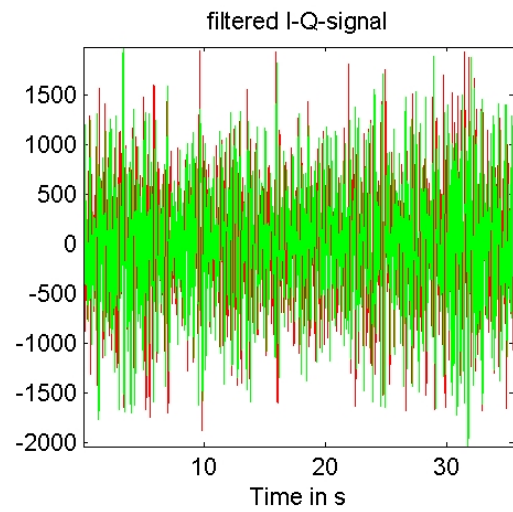
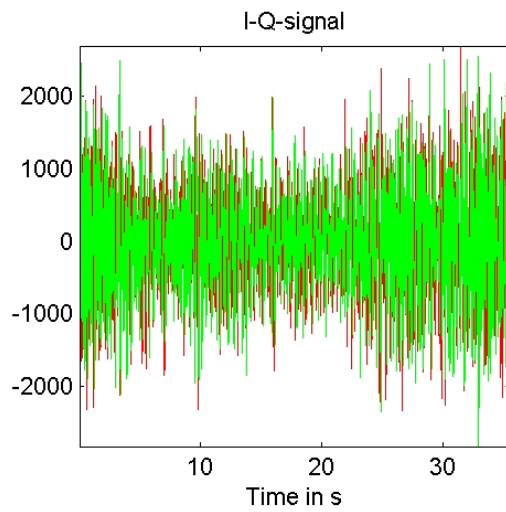


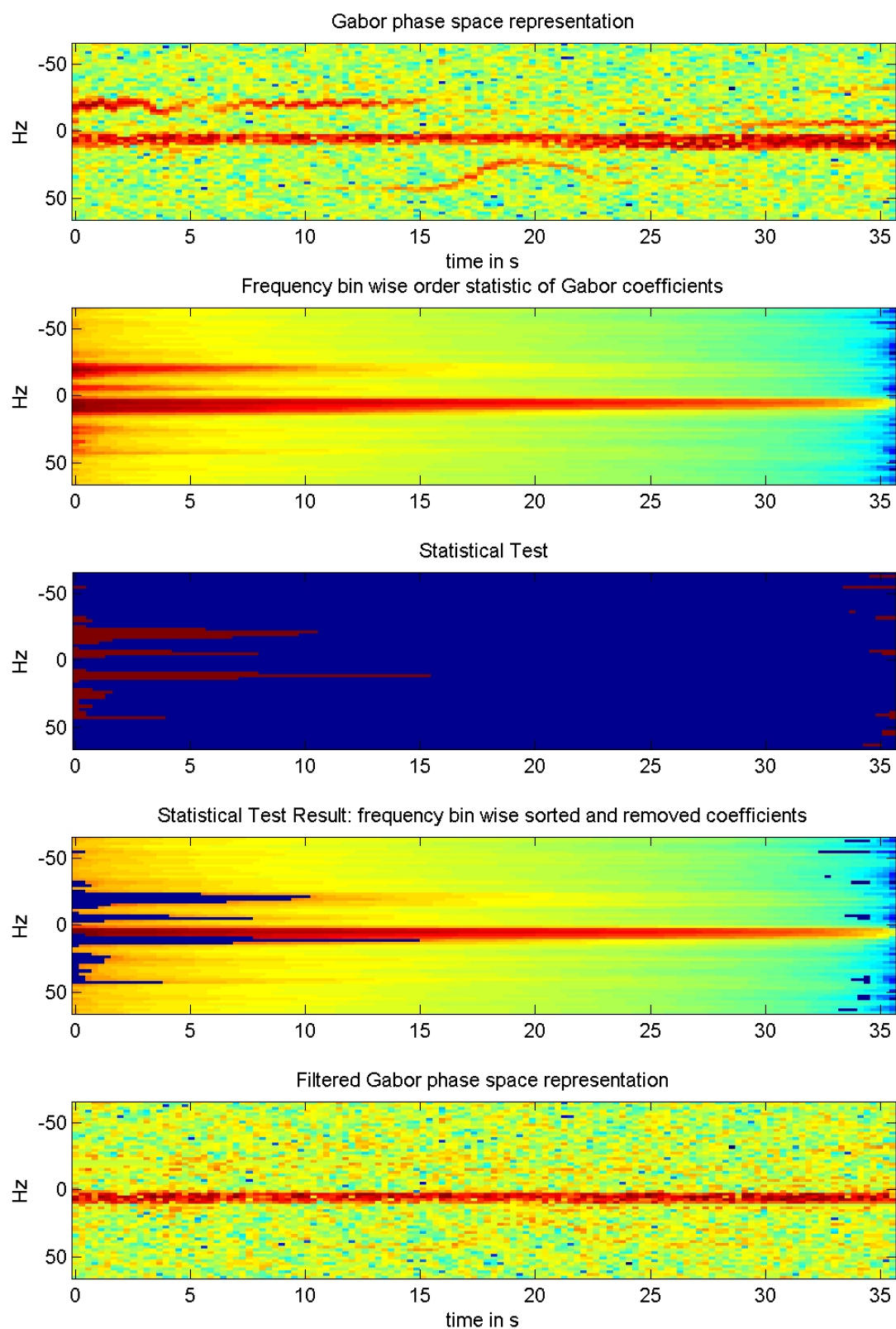


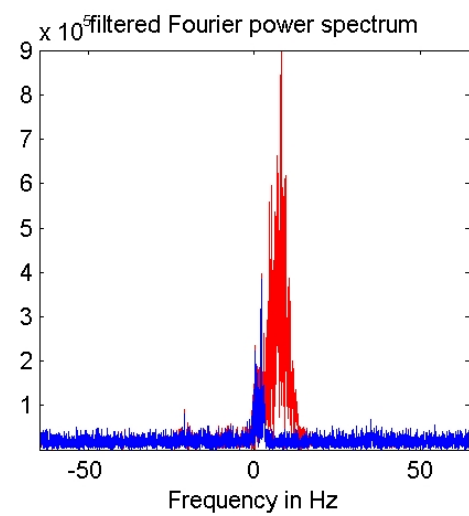
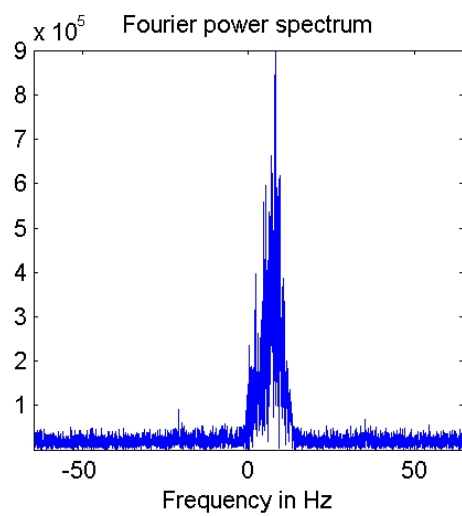
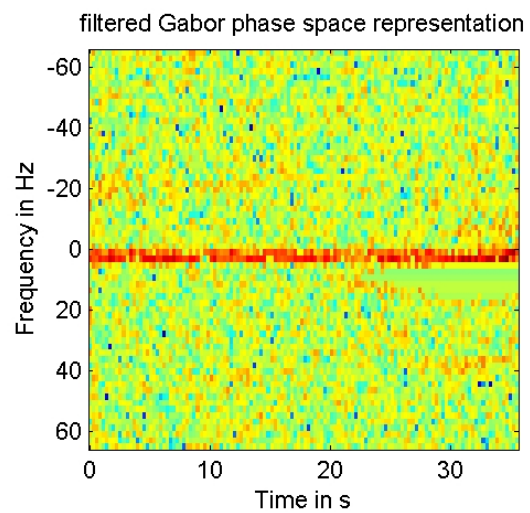
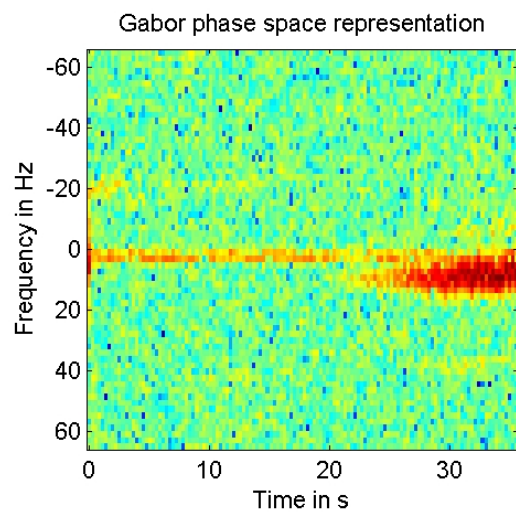
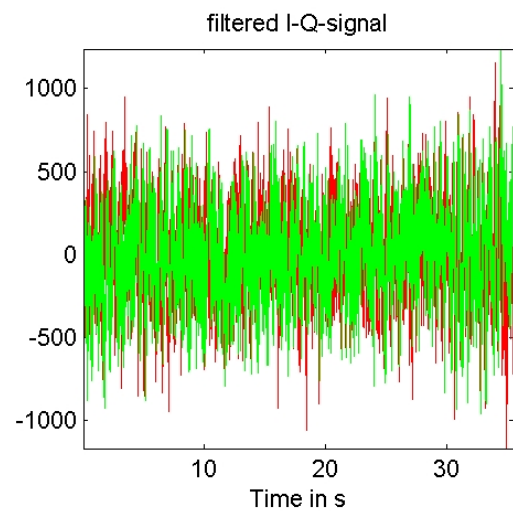
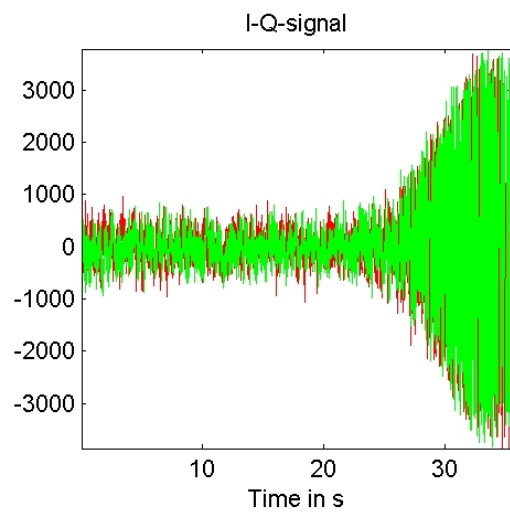
3.4 Dwell 11, Ranges 1–30, Detailed shown range 12, 17, 19

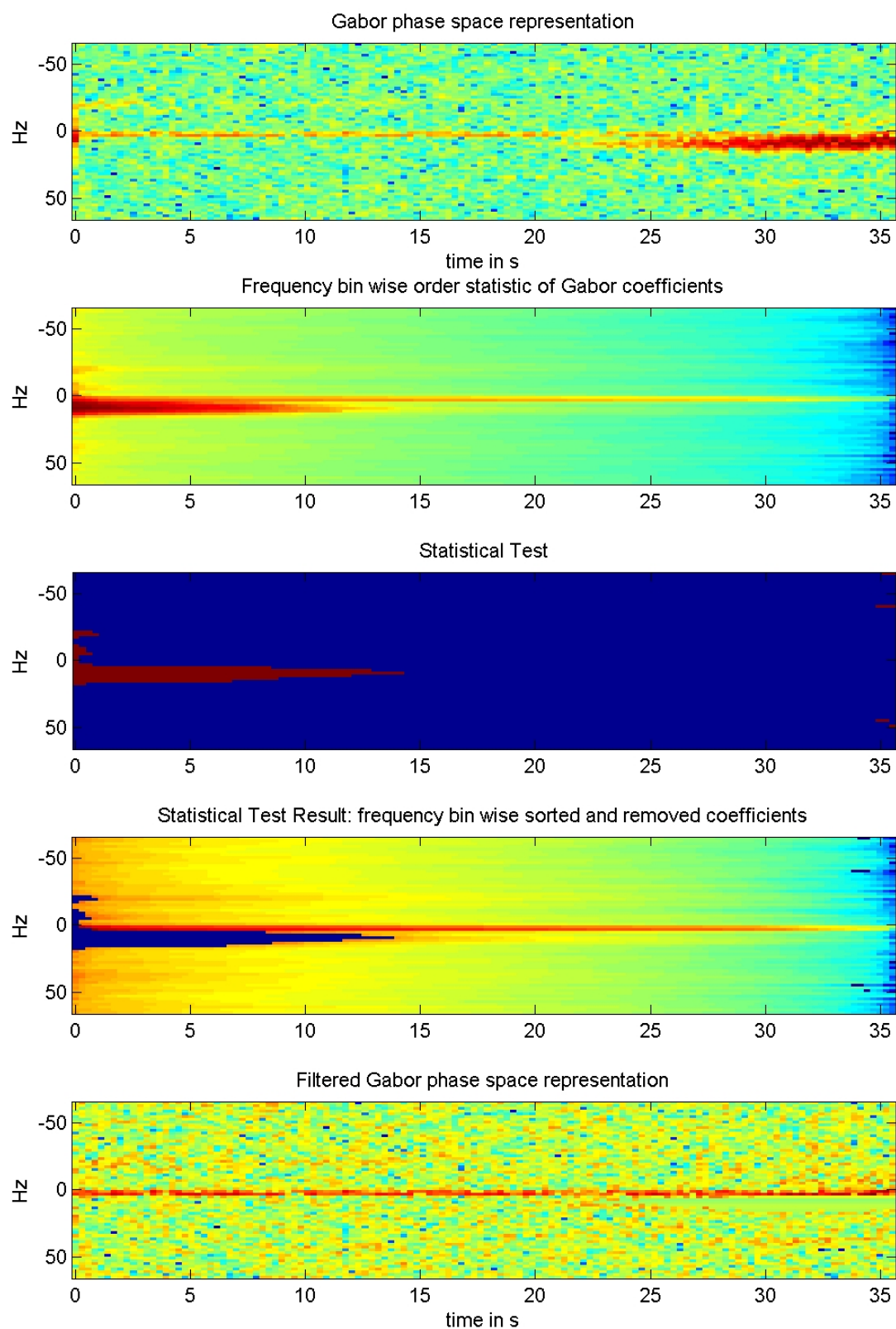


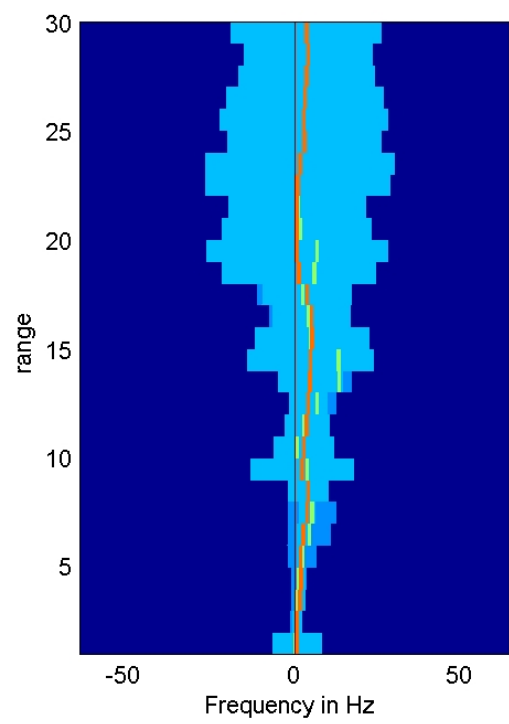
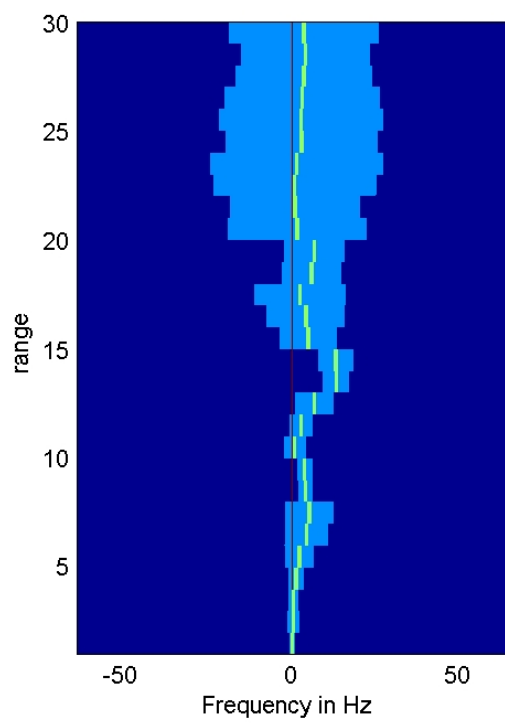
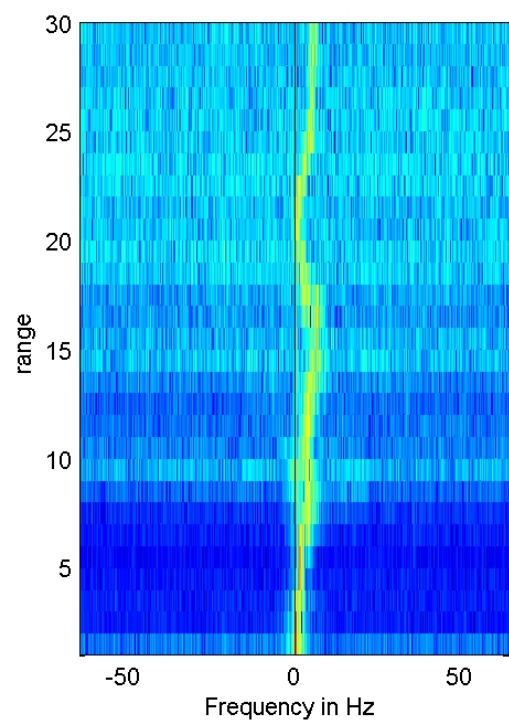
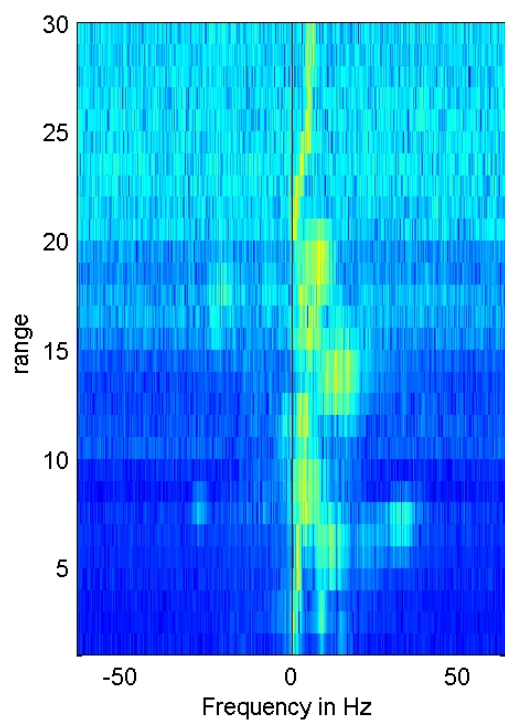












3.5 Prototype: state of the art November 22nd 2005

In the following we give a printout of the prototype code.

- main_DWDdata.m

```
% This program computes the Gabor time frequency representation of
% dwell data; and aims to reconstruct/filter something ...

clear all

%setup of basic parameters:
%=====
filename=['D05286'];
inputpath=['...\Daten\'];
outputpath=['...\Daten\'];
dwells=[10 50];
ranges=[1 30];
pulses=[1 9];
window=['gauss'];s=2; %gauss cos
N=512*(pulses(2)-pulses(1)+1);
delta_t=0.007708; %sampling rate of the I-Q signal
M=128;K=64; %R=M*K/N; redundancy
filter_method=['sam']; %sam row column overall
test_threshold=3; %for the sam test
visualize=['yes']; %yes no
%=====

%creating the analysis and synthesis windows
%=====
g_primal=primal_window(N,window,s);
g_dual=dual_window(g_primal,N,M,K);
%=====

%dwell-wise clutter removal
%=====
for k=dwells(1):dwells(2)
    g_name=[inputpath,filename,'_1dwell_',int2str(k),'.mat'];
    load(g_name,'g');

    for range=ranges(1):ranges(2)
        for pulse=pulses(1):pulses(2)
            f(512*(pulse-1)+1:512*pulse)=g(:,pulse,range);
        end

        [k range]
        [f_rec,freq]=main_tf(f-mean(f),N,M,K,g_primal,g_dual,filter_method,visualize,delta_t,test_threshold);

        %just for spectra plot
        p=2;
        f=abs(fftshift(fft(f)));
        f=f.^p./sum(f.^p);
        f_rec=abs(fftshift(fft(f_rec)));
        f_rec=f_rec.^p./sum(f_rec.^p);
        S(range,:)=f;
        S_rec(range,:)=f_rec;

    end

    %spectra plot
    p=0.1;
    spectra_plot(S,S_rec,p,freq)
    pause
end

end
```

- `function [g]=primal_window(N,atom,s)`
-

```
function [y]=primal_window(N,atom,s)

%
%
% This program computes a vector of lenght N of some preassigned analyzing window
% function.
%
% function [y]=primal_window(N,atom,s)
%
% N- length of signal f
% atom- string determining the window type ('gauss' 'cos')
% s- widht of the atom

s=s*pi/N;

m=0:N-1;
y=0;

if strcmp(atom,'gauss')
    for l=-2:1:2
        y=y+exp(-s*(m+l*N).^2);
    end
end

if strcmp(atom,'cos')
    for l=-2:1:2
        y=y+cos(s*(m+l*N)).*(and(s*(m+l*N)>=-pi/2,s*(m+l*N)<=pi/2));
    end
end

y=y./norm(y);
```

- function [h]=dual_window(g,N,M,K)

```

function [h]=dual_window(g,N,M,K)

%
%
% This program computes a vector of lenght N representing the dual synthesis window
% (with respect to preassigned analyzing window function g).
%
% function [y]=primal_window(g,N,M,K)
%
% g- preassigned analyzing atom
% N- length of signal f
% M- time samples
% K- frequency samples

q=N/M;
p=N/K;

H=zeros(p,M);h=zeros(1,N);b=[1;zeros(p-1,1)];m=(0:M-1);
for n=0:q-1,for t=0:p-1
    A(t+1,:) = K*g(mod(n+t*K+m*q,N)+1);
end;
h(n+m*q+1) = A'*((A*A') \ b);
end

```

- function [f_rec,freq]=main_tf(f,N,M,K,g,h,filter_method,...
visualize,delta_t,test_threshold)

```

function [f_rec,freq]=main_tf(f,N,M,K,g,h,filter_method,visualize,delta_t,test_threshold)

%
%
% This program computes the Gabor time frequency representation (analysis)
% of a given signal f with respect to some analyzing atom window. Moreover,
% it filters the t-f-plane by a hard thresholding routine and reconstructs
% some signal f_rec (synthesis).
%
% A dual (here biorthogonal) synthesis atom is derived automatically.
%
% function [f_rec]=main_tf(f,M,K>window,s,dim,visualize)
%
% f- signal
% N- lenght of f
% M- time samples
% K- frequency samples (for stable analysis/synthesis N<MK)
% window- analyzing atom
% filter_method- thresholding rule
% visualize- just a switch to visualize the output
% delta_t- sampling rate of the signal

time=(1:N)*delta_t;
freq=(-round(N/2)+1:round(N/2))./(N)./delta_t;

% decomposition
F=tf_decomp(f,g,N,M,K);

% some filtering on the data
if strcmp(filter_method,'row')||strcmp(filter_method,'column')||strcmp(filter_method,'overall')
    t=threshold(F,filter_method,K);
    [Fd]=(abs(F)-hard(abs(F),t)).*exp(i*angle(F));
end
if strcmp(filter_method,'sam')
    [Fd]=threshold_sam(abs(F),K,visualize,time,freq,test_threshold).*exp(i*angle(F));
end

% reconstruction
f_rec=tf_reconst(Fd,h);

% visualization of results
if strcmp(visualize,'yes')
    figure(1)
    subplot(3,2,1);
    plot(time,real(f),'r',time,imag(f),'g');axis tight;
    title('I-Q-signal');xlabel('Time in s')
    subplot(3,2,2);
    plot(time,real(f_rec),'r',time,imag(f_rec),'g');axis tight;
    title('filtered I-Q-signal');xlabel('Time in s')
    subplot(3,2,5);
    plot(freq,(abs(fftshift(fft(f)))));axis tight;
    title('Fourier power spectrum');xlabel('Frequency in Hz')
    subplot(3,2,6);
    plot(freq,(abs(fftshift(fft(f_rec))))),'r',...
        freq,(abs(fftshift(fft(f_rec)))),'b');axis tight;
    title('filtered Fourier power spectrum');xlabel('Frequency in Hz')
    subplot(3,2,3);
    imagesc(time,freq,log(abs(F(:,:)))));axis tight;
    title('Gabor phase space representation');ylabel('Frequency in Hz');xlabel('Time in s')
    subplot(3,2,4);
    imagesc(time,freq,log(abs(Fd(:,:)))));
    title('filtered Gabor phase space representation');ylabel('Frequency in Hz');xlabel('Time in s')

    pause
end

```

- function [F]=tf_decomp(f,g,N,M,K)

```
function [F]=tf_decomp(f,g,N,M,K)

%
% This program computes the Gabor time frequency representation (analysis)
% of a given signal f with respect to some analysis atom g.
%
% function [F]=tf_decomp(f,g,N,M,K)
%
% f- signal
% g- analyzing atom
% N- length of f
% M- time samples
% K- frequency samples
%

q=N/M;
p=N/K;

Ff=fft(f);
Fg=fft(g);
F=zeros(K,M);

for k=0:K-1
    Fgf=Fg.*Ff(mod(k*p+(0:N-1),N)+1);
    F(k+1,:)=ifft(sum(reshape(Fgf,M,q),2)/q).';
end;

F = fftshift(F,1);
```


- function [F_rec]=threshold_sam(F,K,visualize,time,freq,test_threshold)

```
function [F_rec]=threshold_sam(F,K,visualize,time,freq,test_threshold)

%Filter function a la Hildebrand&Sekon and Merritt

F_rec=F;
[K,M]=size(F);
[vs,per]=sort(F,2,'descend');
Noise=vs;

for k=1:K

    %sam Test variable computation
    for l=1:M-4
        vv=vs(k,l:M);
        test(l)=mean(vv)^2/var(vv);
    end

    %sam Test
    T(k,:)=(test<test_threshold);
    index_noise=find(test<test_threshold);
    Noise(k,index_noise)=0;
    noise_floor=mean(Noise(k,:)); %(sort_Noise(k,round(M/2))); %the median value

    index=find(test<test_threshold);
    F_rec(k,per(k,index))=noise_floor; %or 0;
end

% visualization of results
if strcmp(visualize,'yes')
    figure(6)
    subplot(5,1,1)
    imagesc(time,freq,log((F)));
    title('Gabor phase space representation');ylabel('Hz');xlabel('time in s')
    subplot(5,1,2)
    imagesc(time,freq,log(vs));
    title('Frequency bin wise order statistic of Gabor coefficients');ylabel('Hz')
    subplot(5,1,3)
    imagesc(time,freq,T);
    title('Statistical Test');ylabel('Hz')
    subplot(5,1,4)
    imagesc(time,freq,(log((Noise))));
    title('Statistical Test Result: frequency bin wise sorted and removed coefficients');
    ylabel('Hz')
    subplot(5,1,5)
    imagesc(time,freq,log((F_rec)));
    title('Filtered Gabor phase space representation');ylabel('Hz');xlabel('time in s')
    pause(0.1)
end
```

- function `f = tf_reconst(F,h)`

```
function f = tf_reconst(F,h)

%
% This program computes the time signal representation (synthesis)
% based on a given Gabor time frequency representation of some
% signal f of length N.
%
% function f = tf_reconst(F,h)
%
% F- Gabor time frequency representation of size K x M (K frequency and M
% time samples)
% h- synthesis window (based on g) of length N

[K,M]=size(F);N=length(h);

q = N/M;
p = N/K;

f=zeros(1,N);

F=fft(fftshift(F,1),[],2);
Fh=fft(h);

for k=0:K-1
    c= repmat(F(k+1,:),1,q) .* Fh;
    f=f+c(mod(-k*p+(0:N-1),N)+1);
end
f = (ifft(f));
```

References

- [1] P. H. Hildebrand and R. S. Sekhon. Objective Determination of the Noise Level in Doppler Spectra. *Journal of Applied Meteorology*, 18:808–812, 1974.
- [2] J. Wexler and S. Raz. Discrete Gabor Expansions. *Signal Processing*, 21:207–220, 1990.