FIFTH INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
DECEMBER 15-18, 1997
ADELAIDE, SOUTH AUSTRALIA

SPECTRAL ESTIMATION ERRORS WHEN USING FFT ANALYZERS

Thomas L. Lagö and Ingvar Claesson

University of Karlskrona/Ronneby
Department of Signal Processing
S-372 25 Ronneby, Sweden

ABSTRACT

An FFT analyzer is often used for spectral estimation. In theory, either a pure sinusoid or white random noise is used. In these two cases it is easy to make an estimate that is correct by using power spectra or power spectral density scaling respectively. In real life, however, signals are likely to be composed of more than one class of signals. This is especially common in the sound and vibration area. Without an a-priori information about the signal, large errors, often several hundred percent are likely. This paper addresses these estimation problems, and covers some of the theoretical background. Easy to use "rules of thumb" are given, that make it possible to verify when a correct estimation has been found. Examples are given using both authentic and synthetic. This type of problem is normally not handled in the text books, and therefore there is a need for raising this issue, especially from a practical point of view, since many engineers are not aware of the problem. From a strict theoretical point of view the problem does not exist, but in reality, if the problem is not dealt with, the consequences can be devastating.

NOTATION

$x_c(t)$	Analog time signal	K_W	Window energy scaling
$x(n)$	Discrete time signal	P_{PS}	Power Spectrum
Δt	Sampling increment	P_{PSD}	Power Spectral Density
T	Time record length	P_{ESD}	Energy Spectral Density
N	Number of samples	BPF	Blade passage frequency
$X_W(f_k)$	Windowed frequency spectra	RPM	Rotation per minute
B_W	FFT analysis bin bandwidth	SPL	Sound Pressure Level

1. INTRODUCTION

Estimation of the amplitude of an unknown signal is one of the basic requirements in measuring techniques. It is common that a-priori knowledge about the signal is given before the measurement takes place. One example is when using a Digital Voltmeter (DVM), and the signal is assumed to be sinusoidal. Another could be when the signal is "broadband" or "noise like". These assumptions have to be fulfilled if the correct amplitude should be estimated.

In the literature for frequency analysis and measurement techniques, a simplification of the signal type is made. It is common to use either the sinusoidal approach or the broadband noise approach. If the signal belongs to one of these classes, several difficulties disappear and some a-priori information about the signal is therefore given. If the signal does not belong to one of these classes, serious absolute measurement errors can result. If a relative measurement are made, the errors are usually small, given that no changes are made on the instrument side. In general, it is impossible to correctly measure a completely unknown signal without the possibility of severe measurement errors. This is in contradiction with most text books, but is what occurs in real life situation, and needs to be treated.

2. MEASUREMENT SETUP

The measurement setup that has been used in the tests is given in Fig. 1 below. The Hewlett-Packard Dynamic Signal Analyzer HP35665A is a common instrument in sound and vibration work, so the setup is typical and many engineers use it every day. A completely unknown signal is fed to the analyzer through a coaxial cable. Despite the fact that the signal is well within the analyzers frequency and dynamic range, negligible aliasing and very good signal conditioning, it is not possible to accurately estimate the amplitude of the unknown signal/signals in the cable. This is difficult for many engineers to understand, but this is a very important fact. Some a-priori information has to be given in order to succeed.

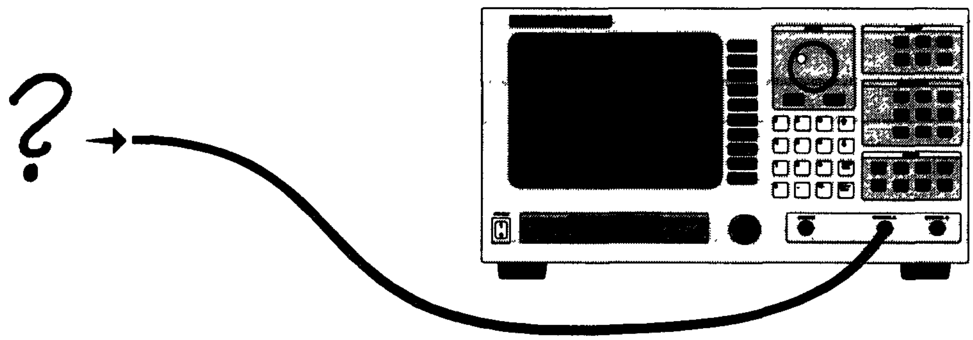

Fig. 1. Illustration of the setup. An HP35665A DSA has been used to collect time data and perform the frequency analysis using an FFT.

When analyzing the unknown signal, the default setup in the analyzer has been used: 400 frequency lines, DC-51.2 kHz frequency span, a Hanning window and one FFT calculation (no average). The results from the measurement using this setup are illustrated in Fig. 2 and Fig. 3.

At first sight, it seems like the signal consists of a sinusoid at 15 kHz with a distortion component at 45 kHz. However, on analyzing the picture in more detail using the markers, it is found that the second peak is located at 46 kHz, not at 45 kHz. Therefore, this peak cannot be a distortion component. Are the first and second peaks really sinusoids? This is very difficult to tell at this stage even though it looks like it on the screen. Therefore, most engineers would have assumed that they are sinusoids. One idea could be to use time domain to see if this gives a clue. The time series will show one sinusoid not two. However, the frequency analysis views two tones not one. By changing window for the frequency analysis the second peak changes its amplitude with more than 30% but not the first peak. This is puzzling. The situation is not good, but real. One could of course avoid this situation by performing only one measurement, reading the markers and be happy. However, if an accurate measurement is the goal, a better understanding of the measurement principles has to be considered.

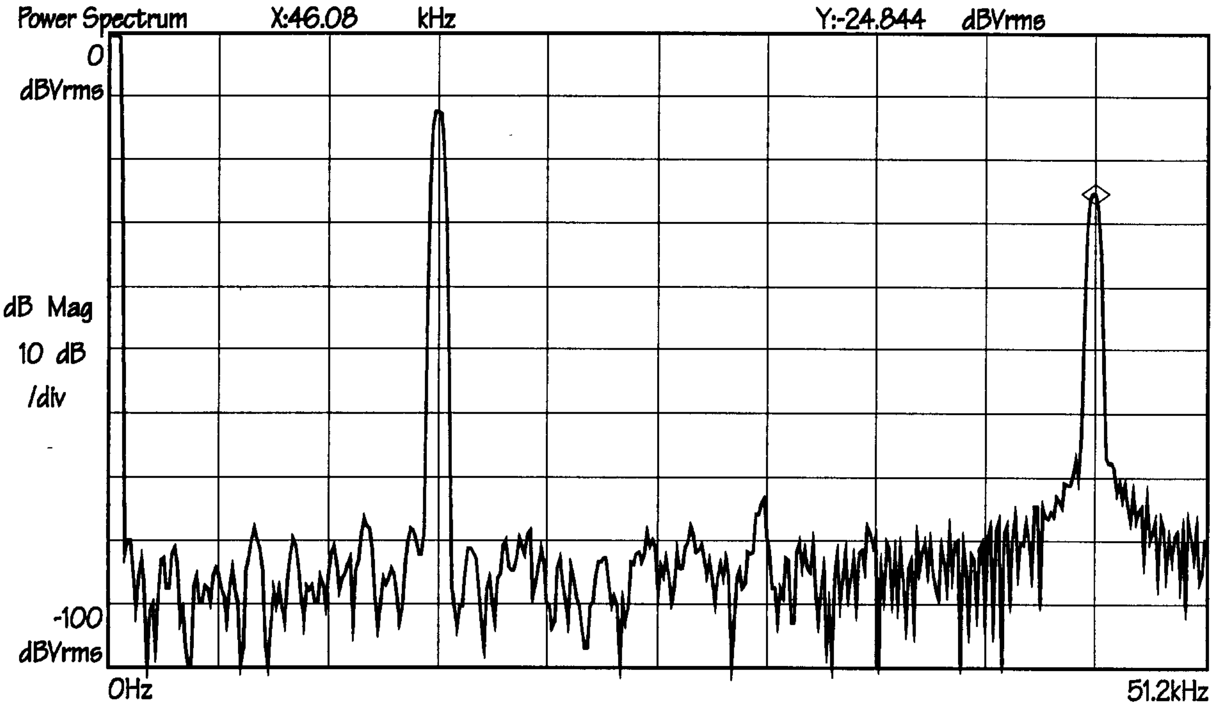

Fig. 2. A frequency analysis on a unknown signal. Two peaks are visible, one at 15 kHz and one at about 45 kHz. There is 12 dB difference in level on the two peaks.

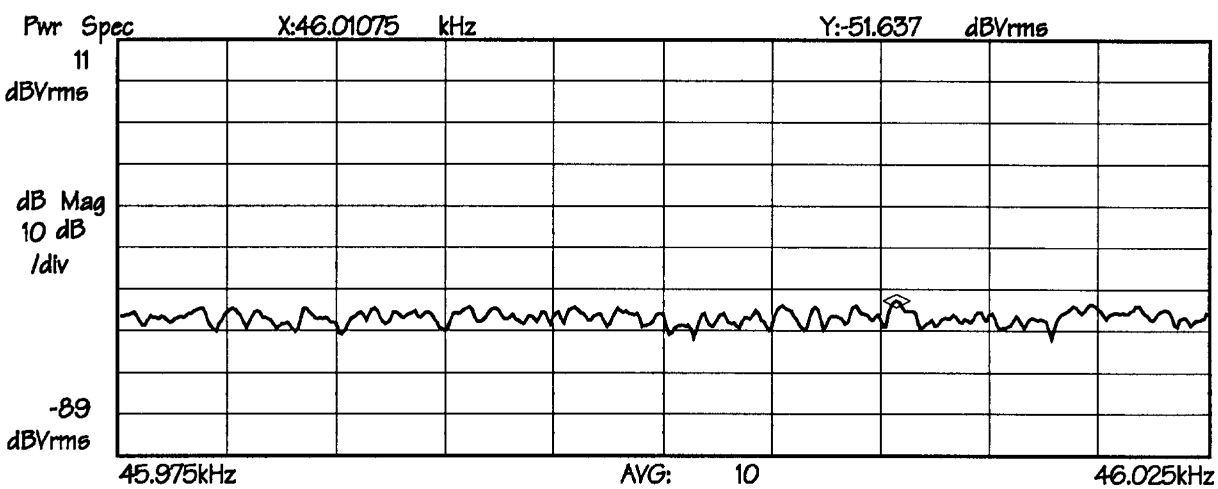

Fig. 3. Frequency analysis of the signal using a zoom technique. The sinusoid is now “white noise” and the level is -52 dB not -25 dB as was given before.

3. AMPLITUDE SCALING

The main reason for the above problem is that the signal at 46 kHz is not a sinusoid, but the signal at 15 kHz is. The 46 kHz component consists of narrowband noise with almost 30 dB lower level than the analyzer shows, given the analysis bandwidth that is used. It is important to note that the problem lies not with the analyzer but the user. The measurement error in this case is 27 dB, which is equivalent to a measurement error of more than 2000%. This problem is due to the fact that the analyzer must be set in a correct scaling method according to the input signal type. The following scaling methods must be used:

- Broadband signals: Use P_{PSD} scaling (Power Spectra)
- Tonal components (sinusoids): Use P_{PS} scaling (Power Density Spectra)
- Transients: Use P_{ESD} scaling (Energy Density Spectra).

If these rules are not followed serious amplitude scaling errors will result, unless an analysis bandwidth of 1 Hz is used. This is how several text books avoid the problem. On the first page they assume that the rest of the book is based on a 1 Hz bandwidth. In this case these scaling problems can be neglected. In real life, it is very rare for the analyzer to use 1 Hz bandwidth. Therefore, this scaling problem must be addressed properly. Some text books state that all signals are "noise like", narrowband or broadband, [1][2][3]. The histogram for a sinusoid will be different from that for a narrowband noise signal. Therefore, the histogram could be a good indication when determining what signal type the signal belongs to. However, most real life signals are composed of several types making it difficult to classify the signal as belonging to one type only. The HP35665 has a facility for performing a histogram calculation, an tool underestimated by many engineers.

4. FREQUENCY ANALYSIS

Assume we have a continuous time series $x_c(t)$ that we wish to sample with equidistant samples, Δt. We will then receive a discrete time series

$$x[n] = x_c(n\Delta t) \qquad n=0,1,2,3,.....,N-1 \tag{1}$$

where N is the number of samples in the series. The corresponding frequency information may be achieved using an FFT, Fast Fourier Transform, which samples the Discrete Fourier Transform, DFT. The frequency information is thus given by

$$X(f_k) = \Delta t \sum_{n=0}^{N-1} x[n] \; e^{-j2\pi f_k n \Delta t} \tag{2}$$

where N is such that the block of data in the transform, is of the length 2^N. The DFT transform will produce frequency information at discrete frequencies given by

$$f_k = \frac{k}{T} = \frac{k}{N\Delta t} \qquad k=0,1,2,3,.....,N-1 \tag{3}$$

where Δt is the sampling increment. If the signal is completely periodic with the length of the time record T, the DFT transform will produce the correct frequency information at the corresponding f_k. If this is not the case, frequency information may leak from one frequency line to another. This leakage effect can be reduced by introducing a time window. A time window $w(n)$ will be multiplied on the time signal $x[n]$ as

$$x_w[n] = x[n] \cdot w[n] \qquad n=0,1,2,3,.....,N-1 \tag{4}$$

and the frequency information will thus be convolved with this window, since a multiplication in time leads to a convolution in frequency,

$$X_w(f_k) = \frac{\Delta t}{K_{w,PS}} \sum_{n=0}^{N-1} x_w[n]\, e^{-j2\pi f_k n \Delta t} \tag{5}$$

where $X_w(f_k)$ denotes the windowed frequency information and $K_{w,\,PS}$ is the amplitude scaling necessary for a sinusoid, due to the decrease in power caused by the window. The window will thus reduce leakage, but also create the analysis bandwidth. There is a trade off between time signal energy, analysis bandwidth, picket fence effect (amplitude ripple), side lobes and spectral leakage. There are several windows available, but the most commonly used in industrial measurements are: No Window (Rectangular), Hanning, Flat top and Exponential. It is important to note that there are several Flat top windows. The P401 by Hewlett Packard has lower side lobes than the P301, which is the flat top window most often used in general measurement equipment. Different windows are presented with their key parameters in **Table 1** below, [4]. It is very clear that the spectral resolution is good with a Hanning window, and amplitude accuracy is best for the flat top window.

Table 1: Description of key window parameters given a frequency range from DC-3.2 kHz, 2048 samples. The bandwidth affects the amplitude scaling.

Window	Amplitude error	Bandwidth, BW	First Sidelobe
Hanning	1.43 dB	6 Hz	-31.5 dB
Hamming	1.75 dB	5.5 Hz	-43.2 dB
Flat top, P301	0.01 dB	13.7 Hz	-70.4 dB
Flat top, P401	0.01 dB	15.3 Hz	-82.1 dB
Rect, no window	3.94 dB	4 Hz	-13.2 dB

The frequency domain signal consists of a real and an imaginary part, and has negative frequencies. In most cases, a Power Spectrum with only positive frequencies is required. This is created by

$$P_{PS} = \begin{cases} \dfrac{1}{(N\Delta t)^2}\, |X_{w,PS}(0)|^2 , & f=0 \\ \dfrac{4}{(N\Delta t)^2}\, |X_{w,PS}(f)|^2 , & f>0 \end{cases} \qquad [V^2] \tag{6}$$

where P_{PS} stands for the Power Spectrum. This amplitude scaling is correct assuming a periodic narrowband signal as the input signal. For white noise, a Power Spectral Density is needed, since this is a broadband signal. The P_{PSD} is given by

$$P_{PSD} = \begin{cases} \dfrac{1}{N\Delta t} \; |X_{w,PSD}(0)|^2 \; , f=0 \\ \dfrac{4}{N\Delta t} \; |X_{w,PSD}(f)|^2 \; , f>0 \end{cases} \qquad \left[\frac{V^2}{Hz}\right] \tag{7}$$

In the above equations, the voltage dimension is included. There are three scaling methods to choose from. This is a difficult choice, since it requires a-priori information about the signal before choosing the right type of scaling. Without this information it is impossible to be sure that the amplitude is correct, [5]. In [3] it is stated that P_{PS} and P_{PSD} are the same. This is a correct statement given the assumption as in the book, of a 1 Hz analysis bandwidth. In most practical cases the analysis bandwidth is not 1 Hz. An amplitude scaling error will thus be the consequence, often of several 1000%.

The above discussion shows that it is important to have a-priori information about the signal in order to evaluate absolute amplitude levels. It is not possible, in general, to find absolute amplitude levels without some knowledge about the signal. This information can be achieved by using a set of measurements and then using that information as the basis for further action.

5. RECOMMENDED MEASUREMENT PROCEDURE

(when absolute amplitude levels are important)

Start the analyzer with a Power Spectrum scaling method (often default). Then,

1. Measure with one frequency range. Read all amplitude levels for all peaks.
2. Measure the same signal again, but with a factor of two decrease in frequency span. This gives twice the measurement time and consequently half the measurement bandwidth B_w. If some amplitude levels (peaks) change levels when compared to the previous measurement, the signal cannot be scaled correctly using P_{PS}.
3. Continue to change the frequency range until the amplitude levels are stable and do not change when the measurement settings are changed. When this happens, the amplitude values can be read, and they have the correct amplitude scaling.

If the signal keeps changing for each change of frequency range, try using P_{PSD} scaling instead. If the levels do not change when using a P_{PSD} scaling, the amplitude levels are correctly scaled. Observe that there are signals where it is not possible to reach a solution for either P_{PS} or P_{PSD}. In such cases, it is difficult to rely on the amplitude values.

A rule of thumb when determining the amplitude scaling method is:

- If the analysis bandwidth « the signal bandwidth: Use P_{PSD} scaling (=broadband signals).
- If the analysis bandwidth » the signal bandwidth: Use P_{PS} scaling (=tonal components).
- If the input signal is transient: Use P_{ESD} scaling (Energy Density) (=transients).

ALWAYS perform <u>two</u> measurements with <u>different</u> analysis bandwidths and compare. If they are equal, the right amplitude scaling is being used!

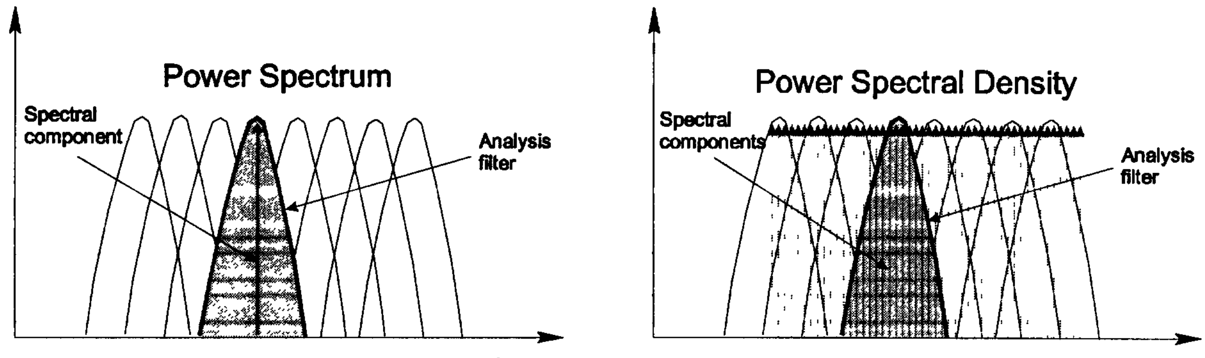

Fig. 4. Illustration of the amplitude situation for a P_{PS} and a P_{PSD} scaling. In the left figure the filter should give the true value, irrespective of the width of the analysis filter. In the right figure, it is necessary to compensate for the width of the analysis filter since the output will be the sum of all components within that filter. This is why a scaling with "Hz" is necessary.

The figure on the left in Fig. 4 illustrates how the scaling is correct using P_{PS}, since there is no compensation for the width of the filter. If there is more than one signal component in the left marked filter, then the amplitude is wrong. The inverse is applicable for the figure on the right. In this case, it is necessary to compensate for the width of the filter. If no compensation is made, the amplitude will be scaled incorrectly. That is why the signal would be overestimated if the analysis bandwidth B_W is larger than 1 Hz, otherwise underestimated.

6. EXAMPLE SIGNALS

Signals from real life situations, especially when dealing with engines, are likely to be composed of combined signal types. In Fig. 5 below, a microphone recording from a Super Puma helicopter has been made. There are several tones which are very close in frequency and this is a complication.

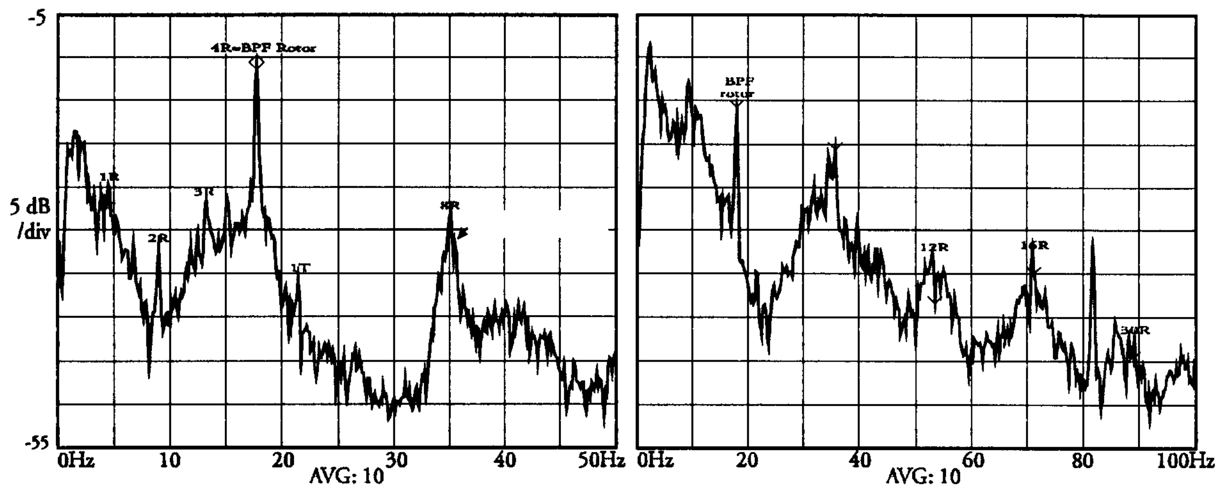

Fig. 5. Illustration of the change in amplitude when a different frequency range is used. The two figures are from two different measurements and can therefore not be compared completely. However, it is obvious that the DC-10 Hz range is overestimated in the figure to the right.

The levels of the different RPMs are important. When using the 100 Hz range it appears that the BPF of the main rotor (17 Hz) is lower than those components below 10 Hz. When the 50 Hz range is used, the reverse is true. By varying the range, different levels of the various RPM components will be found. This must be "wrong." The reason is that each FFT filter is unable to resolve the RPM components that are too close. Therefore, they are "summed" in one FFT filter bin and consequently the levels of the tone/tones are wrong. By following the measurement procedure above, the correct SPL could be found, and this is how it was handled during this measurement, [6]. Without proper treatment of the analysis bandwidth in the above measurement, errors of more than 10 dB would have been the result without proper treatment of the analysis bandwidth.

SUMMARY

Spectral estimation using FFT analyzers such as the HP35665 is very common. If these analyzers are used without a proper knowledge about the analyzed signal, serious amplitude scaling errors may be the result. This has nothing to do with the instrument. Depending on wether the signal is sinusoidal, noise or transient, different scaling methods must be chosen. This requires the user to push a key: Power Spectra, Power Spectral Density or Energy Spectral Density. With the wrong scaling method, errors of several thousand percent can occur. There are also real life signals that are sinusoidal for one component, but broadband for others. In these cases, it is very important to continue adjusting the resolution until the amplitude levels are steady. If not, some components will be estimated with large errors. Several examples from real life measurements illustrate that this is not an academic problem, on the contrary. Most text books avoid the problem by assuming that the analysis bandwidth is 1 Hz whereby the scaling problem does not exist. A recommended analysis technique when performing spectral estimation where amplitude is of importance has been proposed. If this approach is followed, the errors will be controlled.

REFERENCES

1. Bendat J. S. & Piersol A., *Random Data*, New York, 1984.
2. Oppenheim A. V. and Schafer R. W., *Discrete-Time Signal Processing*, Prentice Hall, 1989.
3. Bendat J & Piersol A., *Engineering Applications of Correlation and Spectral Analysis*, Second edition. New York, 1993.
4. Potter R. W., "Compilation of Time Windows and Line Shapes for Fourier Analysis," Hewlett-Packard.
5. Lagö T. L. and Claesson I., "Hur analyserar jag en okänd signal?", SVIB AU-1-konferens, Ronneby, April 1995.
6. Lagö T. L., "Frequency Analysis of Helicopter Sound in the AS332 Super Puma," Research Report, ISSN 1103-1581, ISRN HKR-RES Report 96-8, University of Karlskrona/Ronneby, Sweden, 1996.