FM Analysis/Synthesis Based Audio Coding

Bondhan Winduratna
Laboratorium fuer Informationstechnologie
Universitaet Hannover
Schneiderberg 32, D–30167 Hannover, Germany
windurat@tnt.uni–hannover.de

Abstract

An FM based analysis/synthesis audio coder was developed to encode a signal with a lot number of harmonic components at very low bit rates. The audio signal is modelled by a harmonical line spectrum which is described by FM parameter sets. The estimation of the FM parameters for a given line spectrum is performed by a numerical analysis procedure based on the FM synthesis technique which is well known in music synthesizer engineering. The mean bit rate for the coded FM parameters of single instrument signals varies from 2.1 kbit/s up to 4.8 kbit/s.

1. Introduction

For internet and mobile communications demands for very low bit rate audio coding arise. Audio codecs should be able to represent audio signals at bit rates from 6 to 8 kbit/s. In [1] a survey of audio codecs based on different techniques is given. In the past optimizations of very low bit rate codecs were mainly focused on speech signals . They have very good performance at bit rates from 6 to 24 kbit/s as long as the input material is speech. If they are applied to music signals, it becomes obvious that they are based on a specialized source model which is often not appropriate for general audio signals. On the other hand, transform/subband codecs for high quality audio are available. They show very good performance for coding music signals at bit rates down to 16 kbit/s. The achieved sound quality, however, decreases at lower bit rates, in particular for speech and single instrument signals. In [2] an Audio Analysis/Synthesis Codec (ASAC) is introduced which is capable to transmit music signals at very low bit rates in the range from 6 to 8 kbits. The basic principle of this coder is to analyze the input signal and to encode the most relevant sinusoidal signal components each by two parameters, amplitude and frequency. Since the number of transmitted parameters is limited by the bit rate, the quality of the decoded signal decreases if the input signal contains harmonic tones with a very low fundamental frequency or a large number of harmonic components called partials.

There exists an approach [3] which can be used to overcome the limitations of [2]. In this approach the input signal is analysed with respect to a harmonic line spectrum. The fundamental frequency, amplitudes of line spectrum and number of harmonic are encoded. As illustrated in [3], this approach results in a bit rate of 32kbt/s which is still relatively high for a signal with 25 partials. In order to reduce the required bit rate, a new concept for coding the parameters describing harmonic tones is presented in this paper. It is based on the technique of FM (Frequency Modulation) synthesis which is well known in music synthesizer engineering. The first implementation of this audio coding technique is designed for harmonic tone signals from single instruments.

A short review of FM synthesis is given in Section 2. The FM analysis/synthesis based audio coding is presented in Section 3. The quantizer and coder unit is described in Section 4, and simulation results are shown in Section 5. Finally conclusions are given in Section 6.

2. FM Synthesis

An FM synthesis procedure was created by Chowing [4] to produce a complex audio signal. It's main application was in musical sound synthesis. This FM synthesis offers the advantage, that it is capable to model audio signals which have a large number of partials with relative few parameters. In the simplest FM synthesis, the frequency of a sine wave (carrier) is modulated by another sine wave (modulator) to generate a complex waveform with a spectral characteristics depending on modulation index and the parameters of the two sine waves. The FM signal $x(t)$ is given by (1)

$$x(t) = A \sin[2\pi f_c t + I \sin(2\pi f_m t)] \tag{1}$$

where :

A = amplitude
f_c = carrier frequency
f_m = modulation frequency
I = modulation index
t = time

The fourier series representation of (1) is given by (2)

$$x(t) = \sum_{k=-\infty}^{\infty} J_k(I) sin(2\pi f_c t + 2\pi k f_m t) \tag{2}$$

where the amplitude of each harmonic component of FM spectra depends on the coefficients of the bessel function of the first kind $J_k(I)$, the argument of which is the modulation index. A single FM synthesis process which involves an FM parameter set (modulation frequency, carrier frequency, modulation index and amplitude) is called an FM operator. Several FM operators can be switched in parallel or in series in order to generate more complex spectra. Since there are several possibilities for the FM parameter adjustment, the problem emerges how the FM parameters for a given spectrum should be adjusted correctly. Without assumptions over the given signal characteristics the adjustment for an optimal parameter combination is very difficult and complex. In order to overcome this nonlinear problem of the parameter adjustment, a numeric FM analysis procedure for the estimation of the FM parameters is presented here.

3. FM Analysis/Synthesis based coding

The basic principle of the FM analysis/synthesis based audio coding consists of three steps.

- The first step is exploring the spectral informations of the input signal by using a preanalysis.
- The second step is performing an FM analysis procedure for extracting FM parameter sets from the estimated harmonic spectral lines.
- The third step is quantizing and coding of the estimated FM parameters.

Figure 1 shows the block diagram of this proposed coder.

3.1 Preanalysis

As is well known, most music signals have a harmonical structure. Their spectra are characterized by discrete frequencies which are integer multiples of the fundamental frequency f_0. These informations are packed by the preanalysis into a harmonical spectral lines format as a simplified signal model whose parameters can be extracted by the FM analysis. The preanalysis estimates harmonical spectral lines $\vec{a}$ after [5] from a given input signal. It is started by calculating the relevant spectral lines by usage of a high resolution frequency estimation technique [2] . An iterative search procedure with a cost criterion as presented in (3) is employed to identify the fundamental frequency f_0. The harmonical pattern of integer multiples of f_0 should cover the most relevant spectral lines.

$$G(f) = \sum_n \sum_k \log(a[k])W(f[k], nf) \tag{3}$$

,where

$a[k]$ is amplitude of harmonic component
$f[k]$ is frequency of harmonic component

$$W(f_1,f_2) = \begin{cases} 0 & for\ |f_1 - f_2| > \Delta \\ 1 - \dfrac{|f_1 - f_2|}{\Delta} & for\ |f_1 - f_2| \le \Delta \end{cases}$$

Δ is a threshold value. As an estimate value, the fundamental frequency f_0 is selected, so that $G(f)$ becomes maximal. Subsequently, the partials are assigned. In our implementation, the fundamental frequency f_0 is limited to a minimum of 50 Hz.

3.2. FM Analysis

The task of the FM analysis consists of estimating the FM parameters from a given harmonical spectrum. The basic principle of our FM analysis is to extract FM parameters by estimating the input spectral lines $\vec{a}$ and minimizing the error. The estimated spectral lines are synthesized by means of FM operators which are combined in a parallel structure as shown in figure 2. The optimal FM parameter sets are numerically optimized successively for each FM operator by minimizing the estimation error, i.e. the difference spectrum. A detailed description of the units of the FM analysis is presented below.

Psychoacoustic model

The psychoacoustic model calculates a masked threshold from a given input spectrum. It has been simplified from the model for high quality transform/subband based audio coding as described in [6], the pre– and postmasking effects are not taken into account in this implementation. The masked threshold calculation is started by synthesizing an audio signal from the input spectral lines $\vec{a}$. The next step is a mapping of the complex frequency components of a synthesized audio signal obtained from an FFT based spectral analysis to the partitions with a constant width on the bark scale. After calculating the masked threshold for those partitions they are mapped back to the resolution of the FFT. The masked threshold $\vec{m}$ is used as a criterion. It decides whether a spectrum synthesized from the extracted FM parameter sets sufficiently approximates the input spectrum. If the difference spectrum is below the masked, the FM parameters estimation is stopped.

FM synthesis loop

The FM synthesis loop consists of an FM parameter generator and an FM spectrum synthe–sizer. The first unit generates iteratively the FM parameters (f_m, f_c and I) in all of their possible combinations. Based on these FM parameters, the second unit synthesizes the FM spectrum $\vec{b}$ that actually consists of the bessel coefficients as in the equation (2). The parameters f_m and f_c which are integer multiples of the fundamental frequency f_0, compose the pitch of the FM spectrum. The band width of the FM spectrum depends on the parameter I. In our implementation, the range of I is 0 up to 15,5 with an increment step of 0.5.

Comparator

This comparator has two tasks. The first task is evaluating for each FM operator the optimal FM parameter combination , which is generated iteratively by the previous FM parameter generator. The optimal parameter set is achieved when equation (4) is satisfied, so that the optimization of FM parameters is stopped and the current parameters are selected as an optimal FM parameter set.

$$\sum_{i=1}^{N} |e_n[i] - A_l|b_l[i]|| = Min \tag{4}$$

where,

$e_n[i]$ is the difference spectral line equal to $|a_n[i] - \hat{a}_n[i]|$. $\hat{a}_n[i]$ is the estimated spectral line.

A is the maximum amplitude of the difference spectrum $\vec{e}_n$.

l is the index number of FM parameter set which is optimized in the current process.

n is the frame number

The second task is determining whether the difference spectral lines $\vec{e}_n$ are below the masked threshold $\vec{m}_n$. If one or more components of $\vec{e}_n$ exceed the masked threshold, the first task of the comparator must be executed again until the FM operators fulfill of the condition $\vec{e}_n < \vec{m}_n$.

Accumulator and Predictor

Accumulator and predictor produce the spectrum $\hat{\vec{a}}$ presented in equations (5). The accumulator contributes the first term of equation (5) by accumulating FM spectra whose parameters have been accepted. If a new FM operator is selected, this term is calculated again by adding the new FM spectral lines whose parameters have just been optimized and selected as rth FM parameter set.

$$\hat{\vec{a}}_n = \sum_{r} A_r \vec{b}_r + p_n \hat{\vec{a}}_{n-1} \tag{5}$$

where,

r is the number of FM parameter sets which have been extracted.

The second term of equation (5) is actually a correlation term among consecutive frames. This term is often found in audio signals whose consecutive frames may have similar spectral characteristics. In order to explore this correlation term, a predictor is used if the preanalysis unit detects that the current frame has the same fundamental frequency as the last frame. This second term is calculated by multiplying the estimated spectrum of the last frame $\hat{\vec{a}}_{n-1}$ with a predictor coefficient p_n. The number of FM parameter sets which have to be transmitted should be reduced by using the predictor.

4. Coding of parameters

After estimating the parameters for each frame the next procedure is parameter coding. While the amplitude and fundamental frequency are quantized with a logarithmic scale, the other parameters are quantized with a linear scale. All of the parameters are coded with fix length words as follows

parameter	word length
fundamental frequency f_0	12 bit
carrier frequency f_c	7 bit
modulation frequency f_m	7 bit
modulation index I	5 bit
amplitude factor A	7 bit
predictor coefficient p	8 bit

At the receiver site, the transmitted parameters are decoded and the FM spectrum is synthesized in order to reconstruct the audio signal.

5. Simulation results

For testing the performance of the proposed coder, it has been applied to audio sequences like ”clarinet”, ”horn”, ”german male”, ”german female” etc. at 8kHz sampling rate. Those signals are framed by means of a window function with constant duration of 24 ms.

Table 1 presents the bit rates for the test signals. The mean bit rate varies from 2.1 kbit/s up to 11.9 kbit/s. The reason for this variation is the signal characteristic itself. For signals like horn, clarinet which have a slow temporal change of the signal spectrum, the predictor is effective, it saves more FM operators than for signals like ”german male” which have a strong temporal changing spectrum. In order to evaluate the coding gain of the FM analysis/synthesis procedure, the input line spectrum is directly PCM encoded without FM analysis/synthesis processing. The measurement results for the bit rates in table 1 show that the coding with FM analysis/synthesis reduces the mean bit rate of the coding without FM analysis/synthesis down to 33 % for the signal ”horn”. Furthermore, the peak bit rate which can be achieved by our proposed coder is at about 21.2 kbit/s for the signal ”german male”. For comparison, the peak bit rate which is achieved by parameter based audio coding [2] is at about 45 kbit/s. This value is measured for a harmonic line spectrum with fundamental frequency of 50 Hz and all spectral lines are considered as new individual lines.

Figure 3 presents the bit rates over time (frame). The bit rate shows a maximum where two tones overlap. The preanalysis can only detect the difference frequency of partials from two fundamental frequencies and selects the lowest one as f_0 so that more harmonic components are assigned. Figure 4 shows the spectral lines , the estimated spectral lines and the difference of spectral lines at the end of the FM parameter estimation process for the 20th frame of the signal ”harpsichord”. Figure 5 and 6 show how close is the spectrum of the synthesized signal at decoder to the original one.

6. Conclusions

An FM analysis/synthesis coding is proposed in order to solve the problem of analysis/synthesis coding at very low bit rates associated with signals which have a large number of partials. Applying the proposed concept of FM analysis/synthesis to music and speech signals allows to encode a group of up to 32 partials by only 6 parameters instead of 34 parameters. The simulation results show that the FM analysis/synthesis can reduce the mean bit rate down to 33 % when compared to PCM coding of the amplitudes of the partials.

The computational cost for an encoder based on the proposed technique is relatively high. Therefore, an acceleration procedure must be developed in order to enable a real time encoding. Research on fundamental frequency estimation must be carried on to improve its accuracy which is an important factor to overcome the problems of this coder on multi tone signals.

This work has been done under contract of and in close cooperation with the Deutsche Telekom Berkom GmbH, Berlin, Germany.

7. References :

[1] Edler, B. "Very Low Bit Rate Audio Coding Development", AES 14th International Conference, Juni 1997

[2] Edler, B.,Purnhagen, H. and Ferekidis, C. "ASAC – Analysis/Synthesis Codec for very low Bit Rates", Preprint 4179, 100th AES Convention, Copenhagen, May 1996

[3] Ferreira,A.J.S. "Perceptual Coding of Harmonic Signals", Preprint 4177, 100th AES Convention, Copenhagen, May 1996

[4] Chowing J. "The Synthesis of Complex Audio Spectra by Means of Frequency Modula tion", Journal of The Audio Engineering Society, Vol. 21, Number 7, 526–529, September 1973.

[5] Kodate, A. "Fundamental Frequency estimation", interner Abschlußbericht, Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Oktober 1995.

[6] Fuchs, H., Baumgarte, F., and Ferekidis, C. "A Nonlinear Psychoacoustic Model Applied to the ISO MPEG Layer 3 Coder", Preprint 4087, 99th AES Convention, New York, October 1995

[7] Winduratna, B. "FM–Analyse/Synthese von Audiosignalen", Studienbericht im Rahmen des DAAD–Programs, Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Mai 1995.

[8] Hesse H., "Parametrisierung der Spektren von Einzelinstrumentsignalen fuer den Einsatz in einem Analyse/Synthese–Coder", Diplomarbeit am Institut fuer Theore–tische Nachrichtentechnik, Universitaet Hannover, Dezember 1995

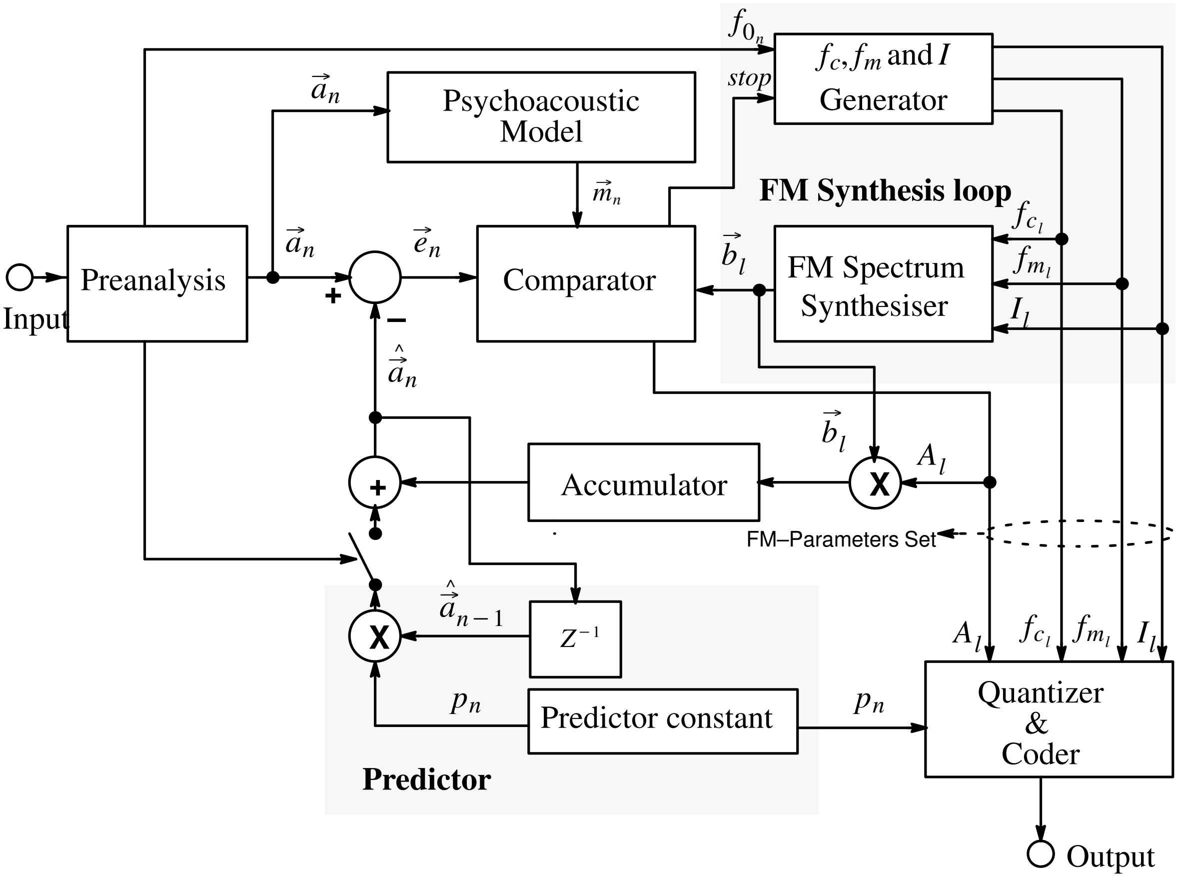

Figure 1 Block diagram of analysis/synthesis based audio coding

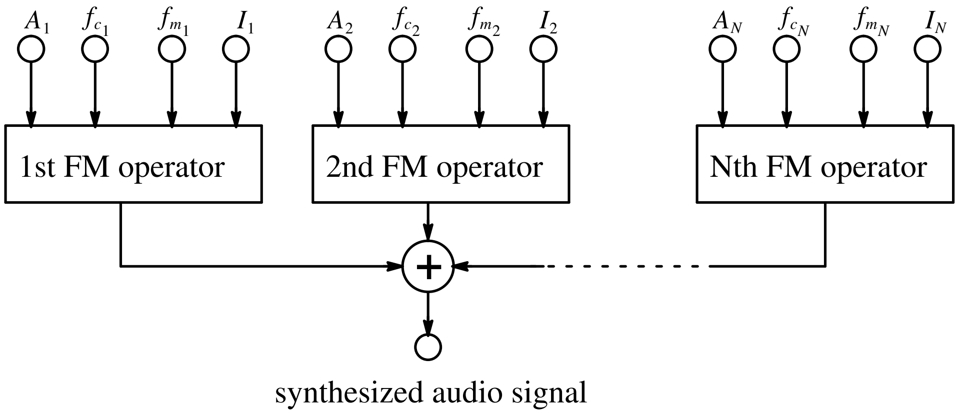

Figure 2 N FM operators are combined in the parallel structure

audio signal	maximum bit rate with FM	mean bit rate with FM	maximum bit rate without FM	mean bit rate without FM
Clarinet	12.9 kbit/s	3.3 kbit/s	19.7 kbit/s	5,2 kbit/s
Harpsichord	17.5 kbit/s	4.5 kbit/s	29.2 kbit/s	7,4 kbit/s
Horn	9.1 kbit/s	2.1 kbit/s	17.6 kbit/s	6,3 kbit/s
Trumpet	16.3 kbit/s	4.8 kbit/s	24.7 kbit/s	5,3 kbit/s
German Female	21.2 kbit/s	11.2 kbit/s	29.2 kbit/s	13,9 kbit/s
German Male	21.2 kbit/s	11.9 kbit/s	29.2 kbit/s	14,8 kbit/s

Table 1 The bit rate measurements from the FM analysis/synthesis based audio coding.

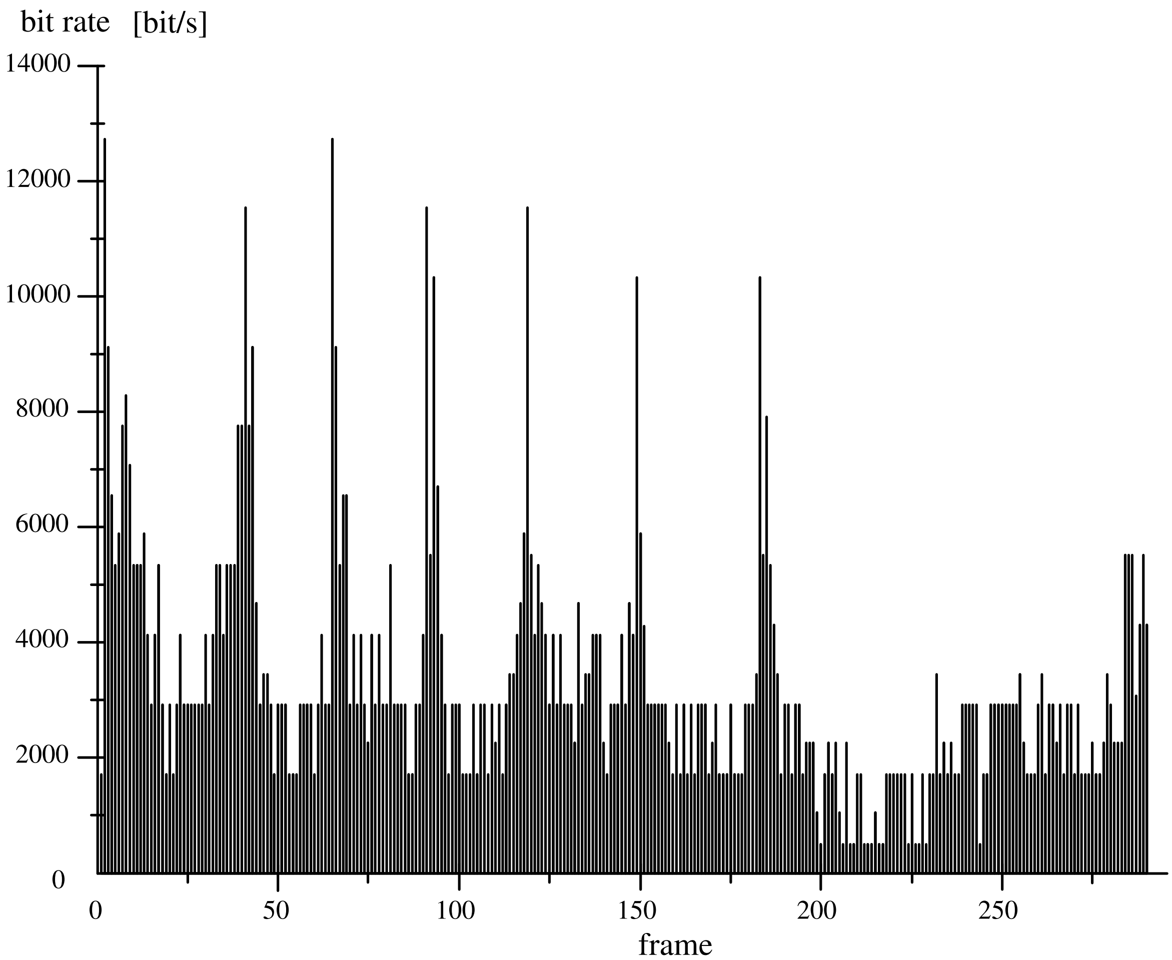

Figure 3 The necessary bit rates of the signal ”clarinet”.

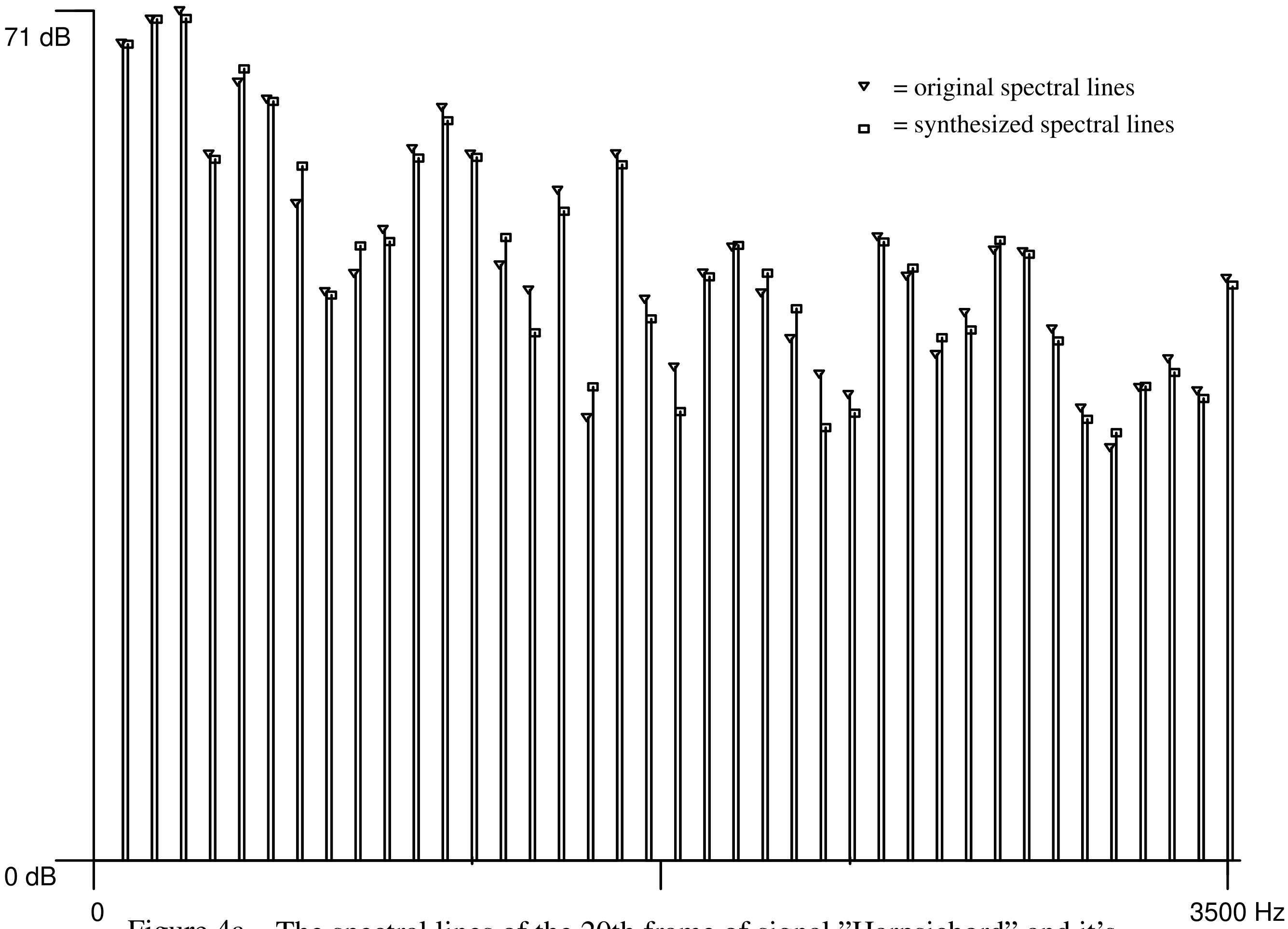

Figure 4a The spectral lines of the 20th frame of signal "Harpsichord" and it's synthesised spectral lines.

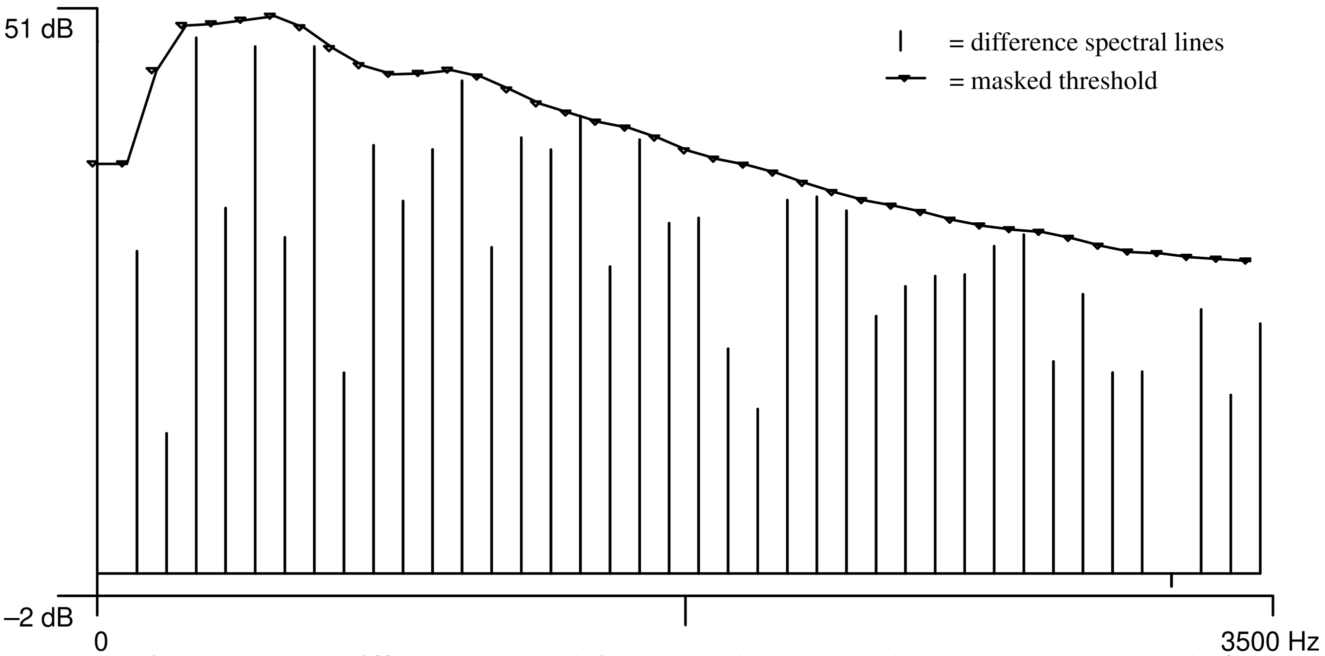

Figure 4b The difference spectral lines are below the masked threshold at the end of the FM analysis process for the 20th frame of signal "Harpsichord".

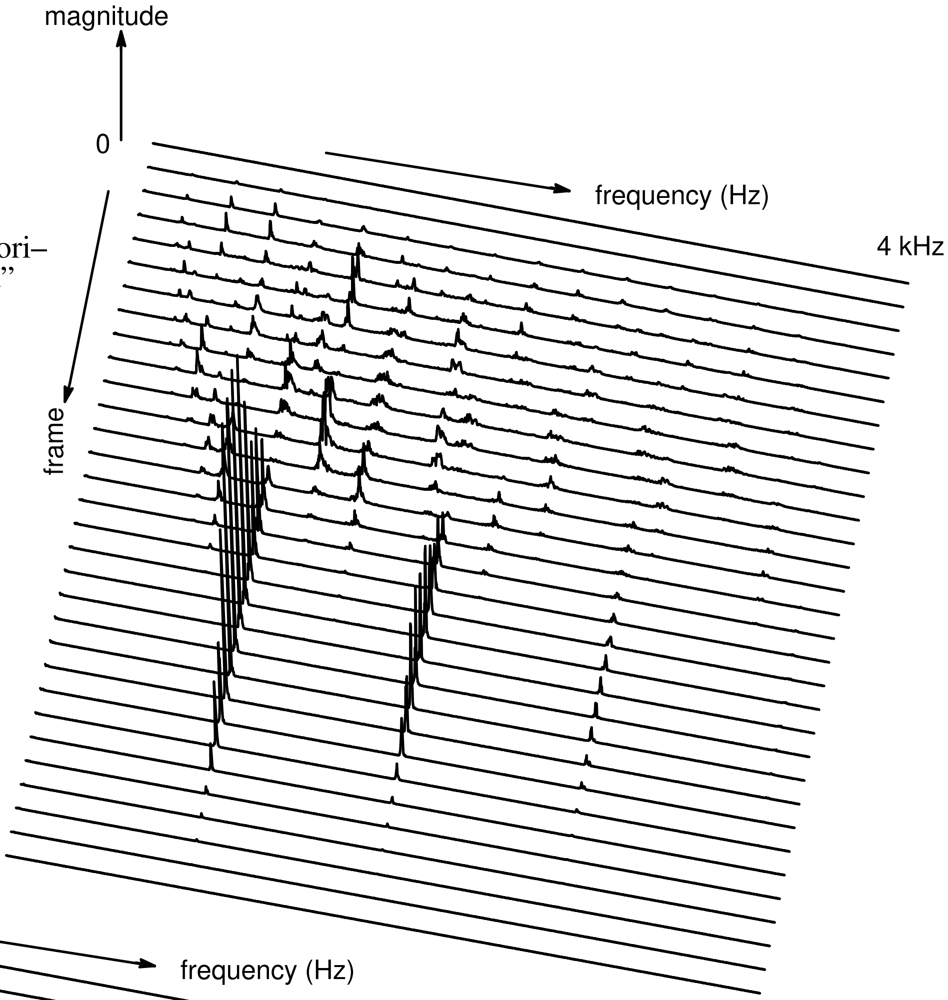

Figure 5a The spectrum of the ori– ginal signal ”trumpet” at the encoder input

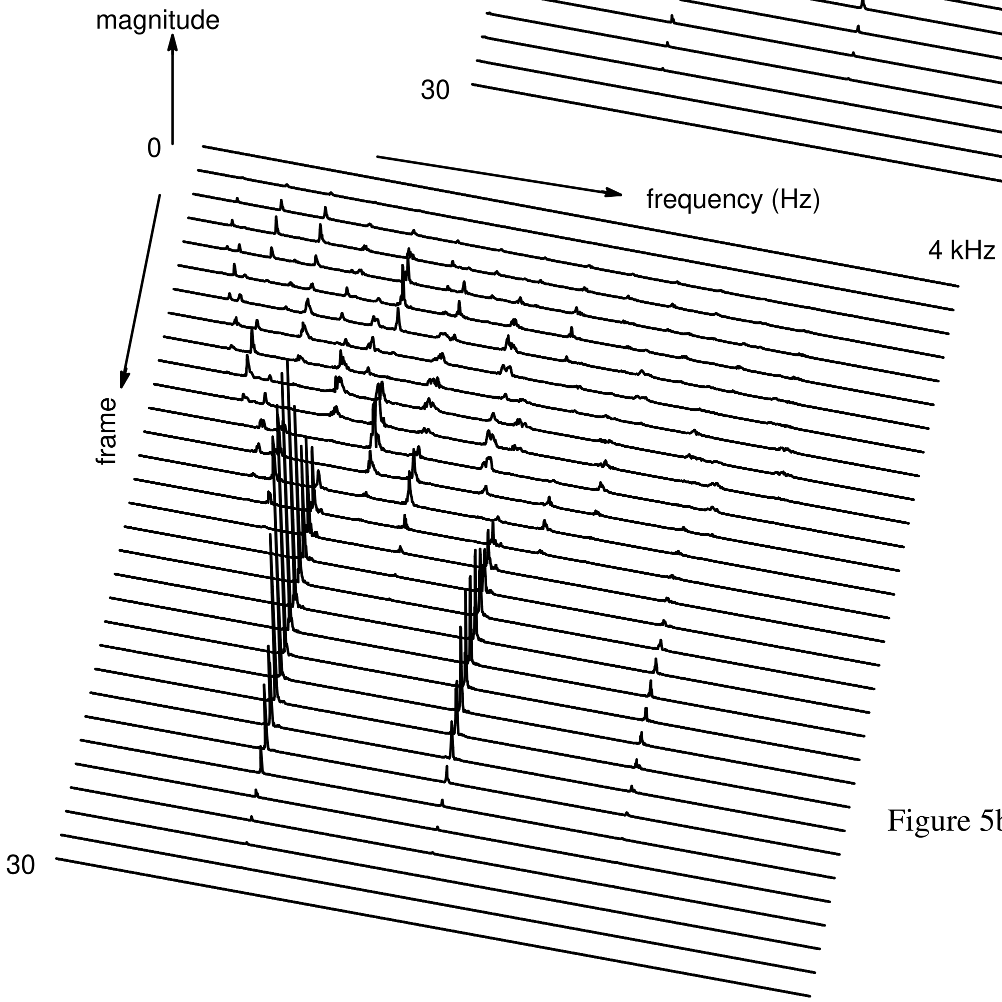

Figure 5b The spectrum of the synthe– sized signal ”trumpet” at the decoder output.

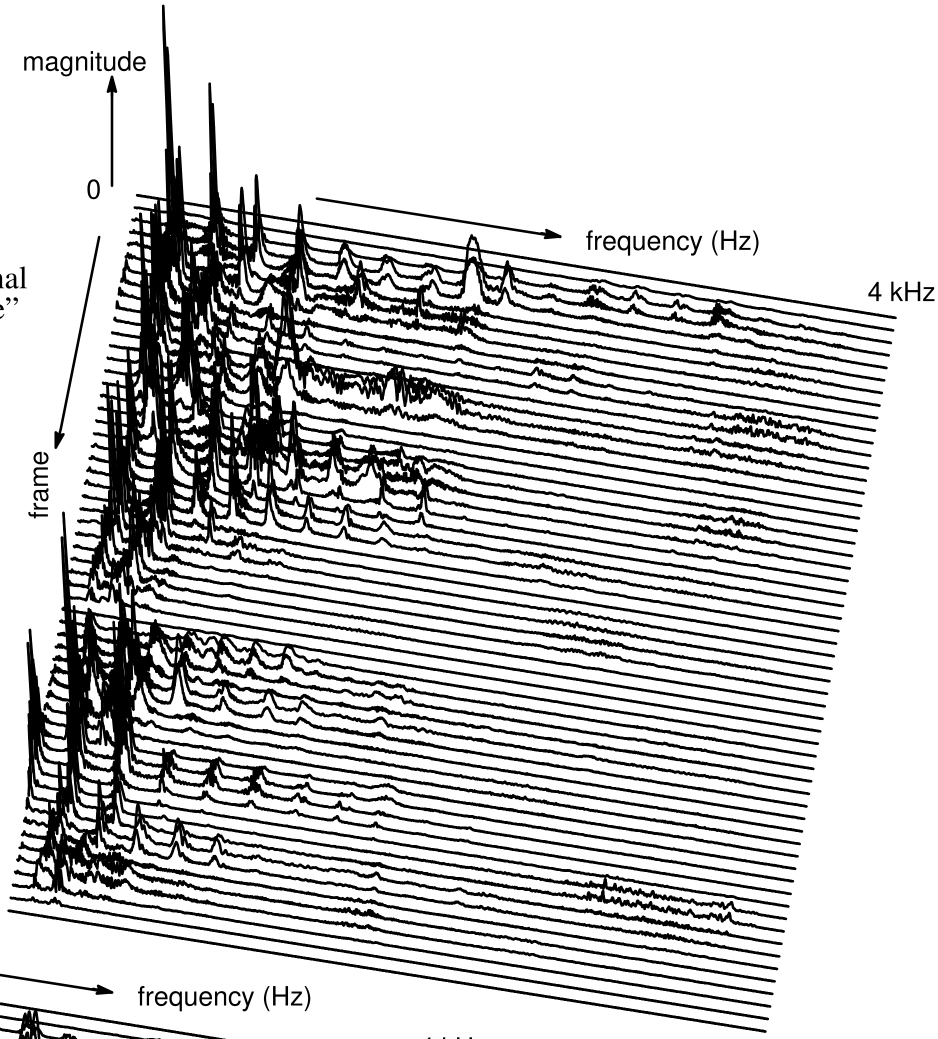

Figure 6a Spectrum of the original signal "german female" at the encoder input

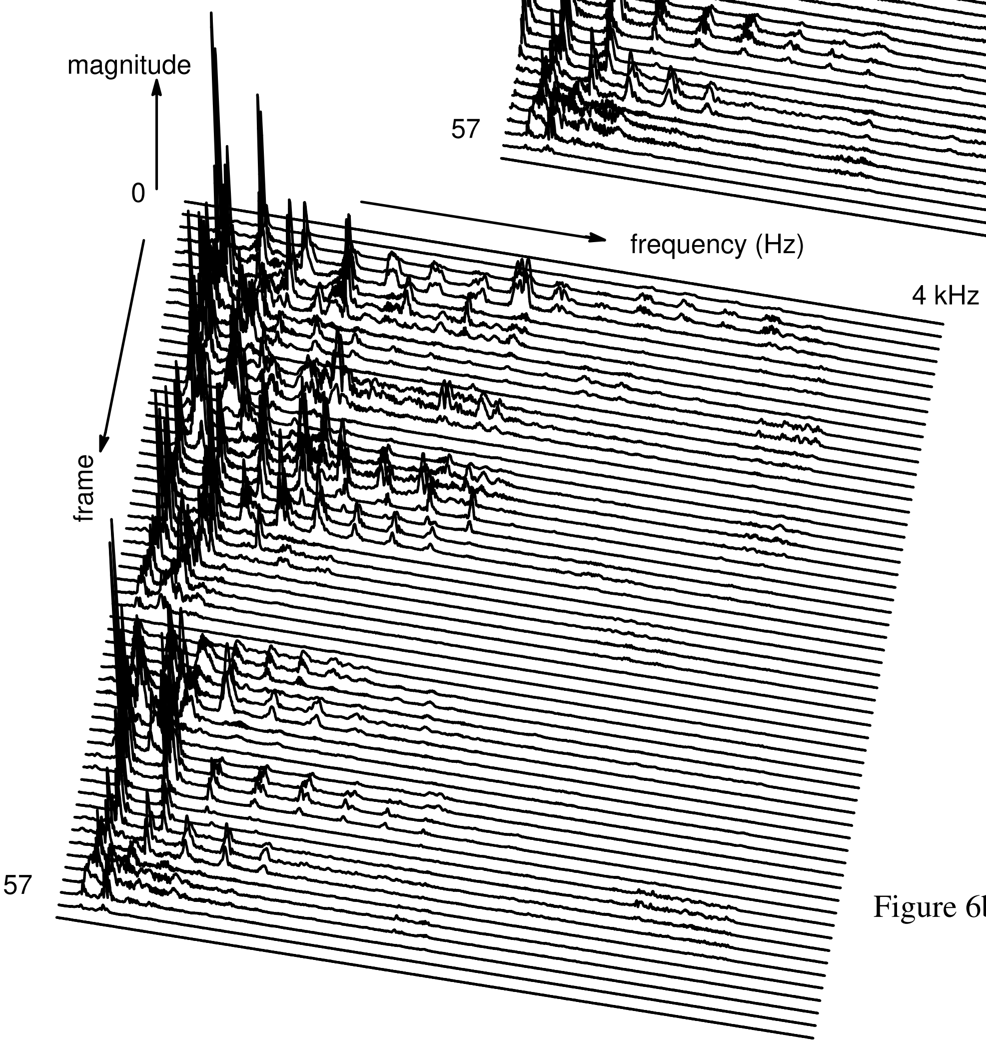

Figure 6b Spectrum of the synthesized signal "german female" at the decoder output.