

Computer Processing of Audio Signals by Exclusion Filters.

Robert C. Maher

Dept. of Electrical Engineering
Univ. of Nebraska-Lincoln
209N WSEC
Lincoln, NE 68588-0511
(402) 472-3771
eerdrcm@engvms.unl.edu

(PACS) Subject Classification Numbers: 43.75.Wx 43.85.Ta

Abstract

A common task in computer music and electroacoustic signal processing is additive *mixing* of two audio signals. If the two input signals contain discrete spectral components, their sum will typically contain amplitude beating and other interactions between pairs of components with similar frequencies. A new method is described which suppresses spectral interactions during mixing by deriving a time variant "exclusion filter" from the short-time spectrum of one of the signals in order to prefilter the other signal. This technique allows one of the signals (dominant) to pass through the mixing process with little modification, while the other signal (secondary) is prevented from interaction by attenuation of its conflicting spectral components. The exclusion filter is specified in a flexible manner, which can include such psychoacoustic criteria as critical bands.

Introduction

The presence of *beats* due to nearly coincident frequency components is a well-known phenomenon in musical acoustics. In live performance situations with adjustable pitch instruments, skilled musicians may adjust the fundamental frequency of each note in order to minimize beats from one note to the next according to the musical context or to match other performers in an ensemble [Benade, 1976]. However, these adjustments are difficult or impossible when dealing with fixed pitch instruments (keyboards, chimes, etc.) or when additively mixing prerecorded material in commercial recording functions. The research reported here is a preliminary investigation of digital signal processing methods for controlling component beating occurring when audio signals are added together. This control includes both enhancement or suppression of beating according to the needs of each particular recording situation.

In the audio *mixing* process, two or more input signals are additively combined to create a composite output signal.* The input signals can be weighted (amplified or attenuated) prior to being summed in order to adjust the relative contribution of each input to the total output signal. If two spectral components with similar frequencies are present in the output signal, amplitude beats will be produced according to the standard relationship:

$$A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t) = [A_1 + A_2] \cdot \cos(\omega_a t) \cdot \cos(\omega_d t) - [A_1 - A_2] \cdot \sin(\omega_a t) \cdot \sin(\omega_d t), \quad (1a)$$

which can be expressed in polar form as

$$\sqrt{A_1^2 + A_2^2 + 2 A_1 A_2 \cos(2 \omega_d t)} \cdot \cos(\omega_a t + \arctan[\tan(\omega_d t) \cdot (A_1 - A_2)/(A_1 + A_2)]) \quad (1b)$$

where $\omega_d = (\omega_1 - \omega_2)/2$, and $\omega_a = (\omega_1 + \omega_2)/2$.

Several examples of beats for combinations of two sinusoidal input signals are given in **Figure 1**.

In common musical situations involving pitched (voiced) signals, the various quasi-harmonic partials of each musical voice will *collide* with components from the other voices, resulting in various combinations of beats which vary as the partial frequencies change from note to note or because of vibrato. The amplitude of the various spectral components varies with time also, resulting in much more complex beating patterns than indicated by Equation 1b. Further, interesting monaural and binaural processes within the human auditory system can create beating effects [cf., for example, Stevens and Davis, 1983]. Depending upon the disposition of the listener, the presence of beats can either be interpreted as an unwanted corruption of the "true" musical message or as a desirable, natural way in which two different musical lines interact. In either case we may wish to emphasize or deemphasize the beats via electronic means in the manner of skilled musicians in a live performance.

Sinusoidal Analysis

The analysis method for this investigation uses a sinusoidal representation which has been found to be useful for speech, music, bioacoustic sounds, etc. [McAulay and Quatieri, 1986; Maher and Beauchamp, 1990; Maher, 1990]. The MQ (for McAulay and

* Note the difference between *additive mixing* assumed here and *multiplicative mixing* often assumed in communication circuits.

Quatieri) representation can be considered a generalization of simple Fourier analysis to include time-variant spectra and possibly non-harmonic partials. The analysis procedure begins by dividing the digitized input signal into overlapping sections called *frames*. Each frame is multiplied ("windowed") by a lowpass window function to reduce spectral leakage, followed by calculation of the Discrete Fourier Transform (DFT) using a zero-padded Fast Fourier Transform (FFT) algorithm. The magnitude of the DFT is computed and all "peaks" in the magnitude spectrum are identified and attributed to *underlying sinusoidal components* at those frequencies. The process is repeated for each input frame and the spectral peak information (amplitude, frequency, and phase) is matched from frame to frame in order to follow changes in the input signal. The input signal can then be regenerated by an additive synthesis procedure using the amplitude and frequency information obtained for each frame and a smoothly interpolated phase function. This analysis/synthesis process does not necessarily form an identity system, but the resynthesis results have been found to be very good for many types of input signals [Serra, 1989], although this has not yet been subjected to any formal psychoacoustic testing.

The MQ analysis process is performed on each of the input signals. The lists of discrete spectral components (peaks) for corresponding frames of each input signal are then compared: if a component frequency from one signal is within some preselected *collision range*, denoted Δf , of any components in the other signals, amplitude beating between these components in the mixed output is indicated. Since the amplitude and frequency of each component have been identified during the analysis the collision range detection could also include comparisons of the component amplitudes and frequencies in terms of empirical psychoacoustic information, such as critical bands. This line of research will be investigated more fully in the future.

Once the beating components are identified using the chosen collision range criterion several strategies can be employed to alter the extent of the amplitude beats. The strategy considered in this research is an *exclusion filter*, $E(f)$, which selectively reduces the amplitude of interfering (beating) partials of one or more of the input signals while allowing one of the input signals to pass unaltered. The unaltered signal will be referred to as the *dominant* signal, with the other signals being *secondary*. For example, the dominant signal could be a particular musical line which is desired to stand out from the rest, such as a vocal part, while the secondary signal(s) could be background material or accompaniment. In this investigation the exclusion filter is actually better described as a *weighting mask applied to the sinusoidal analysis data of each secondary voice prior to resynthesis*. Two elementary example forms for the exclusion filter are:

$$E(f) = \begin{cases} 1 & , \quad \text{for } |(f - f_i)| > \Delta f \\ \left[\frac{(f - f_i)}{\Delta f} \right]^2 & , \quad \text{for } |(f - f_i)| < \Delta f \end{cases} \quad (2)$$

or

$$E(f) = \begin{cases} 1 & , \quad \text{for } |(f - f_i)| > \delta \cdot f_i \\ \left[\frac{(f - f_i)}{\delta \cdot f_i} \right]^2 & , \quad \text{for } |(f - f_i)| < \delta \cdot f_i \end{cases} \quad (3)$$

where the f_i ($i=1,2,3,\dots$) are the component frequencies (partials) determined by the analysis of the dominant signal, and δ in **Equation 3** is a fractional parameter specifying the bandwidth of the exclusion filter as a function of each dominant component frequency. The second-order (parabolic) dependence of the exclusion filter attenuation upon the frequency difference $f - f_i$ was chosen to avoid an abrupt transition between the attenuation and no-attenuation ranges. Note also that the exclusion filter described by **Equation 2** uses a fixed value of the collision range, Δf , while **Equation 3** uses a frequency dependent range, $\delta \cdot f_i$. The fixed range provides beat reduction based upon some minimum allowable beat frequency, say, 12 Hz, while the frequency dependent range diminishes beats in a constant-Q fashion under the hypothesis that this reduces the perceptual "roughness" encountered when two high frequency components are present within one critical band. Extensions of these simple exclusion filters to include amplitude comparisons between partials are also under consideration in this research.

At present, this investigation has concentrated on the use of the exclusion filter to attenuate any components from other signals in the vicinity of the dominant signal's partials by adjusting the sinusoidal analysis data for each of the input signals. This approach has the significant drawback that *each* of the input signals must be passed through the sinusoidal analysis process, requiring substantial processing overhead. Another approach might be to construct a digital filter with adjustable zeros corresponding to the partial frequencies of the dominant signal. The zero locations would be determined from a sinusoidal analysis of the dominant signal, while the secondary signals would be passed through the digital filter without being analyzed separately .

Examples

Two simple examples are given to demonstrate the exclusion filter process. These synthesized examples are not intended to be comprehensive, rather they exhibit the main features of this preliminary research.

In example 1, two input signals are present. Both contain five equal-amplitude harmonic partials and are 8 seconds in duration. Signal 1 has a constant fundamental frequency of 261.6 Hz (middle C), while Signal 2 has a varying fundamental frequency between 200 Hz and 256 Hz. Signal 1 was arbitrarily chosen to be dominant in this example. The signal *envelope* for a portion of the *unprocessed* sum of the two input signals is shown in **Figure 2a**. The same segment for the *processed* sum of the two input signals using a simple exclusion filter for beat reduction is shown in **Figure 2b**. Note the reduction in intercomponent beating. **Figure 2c** shows the MQ analysis of the *processed* mix.

A more complex case is considered for example 2. In this case one input signal consists of five constant frequency, equal-amplitude harmonic partials (261.63 Hz fundamental), while the other input also contains five equal-amplitude harmonic partials but with fundamental frequency steps corresponding to the equal-tempered pitches E, G, and F above middle C, forming intervals of major 3rd, perfect 5th and perfect 4th. In this example the constant frequency voice was again (arbitrarily) assigned to have precedence over the varying frequency voice, resulting in the exclusion filter indicated by the curve in **Figure 3a**. **Figure 3b** shows the overlapped spectra following the exclusion filter process for the pitch combination C:G. Note the conversion of the equal-tempered intervals by suppression of beating from the conflicting partials. The MQ analysis of the processed mix is shown in **Figure 3c**.

Conclusion

Methods for time-variant manipulation of spectral components show great promise for future work in electroacoustic signal processing. The increasing use of digital processing systems during music recording and mastering indicates an expanding role for new musical manipulation concepts. In particular, the use of psychoacoustic criteria for coding and processing audio material is becoming a practical possibility in digital audio engineering.

References

Benade, A. H. (1976), *Fundamentals of Musical Acoustics*, New York: Oxford University Press

McAulay, R. J. and Quatieri, T. F. (1986), "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 4

Maher, R. C. (1990), "Evaluation of a method for separating digitized duet signals," *in press* (to be published Dec., 1990), *J. Audio Eng. Soc.*

Maher, R. C. and Beauchamp, J. W. (1990), "An investigation of vocal vibrato for synthesis," *Applied Acoustics*, vol. 30, no. 2-3, pp. 219-245

Serra, X. (1989), "A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition," Ph.D. Dissertation, Stanford University, Stanford, CA

Stevens, S. S., and Davis, H. (1983), *Hearing: Its Psychology and Physiology*, New York: American Institute of Physics (originally published 1938)

Figure 1: Amplitude beating of sinusoidal components.

$$A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t) = [A_1 + A_2] \cdot \cos(\omega_a t) \cdot \cos(\omega_d t) - [A_1 - A_2] \cdot \sin(\omega_a t) \cdot \sin(\omega_d t)$$

which can be expressed in polar form as:

$$\sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos(2\omega_d t)} \cdot \cos(\omega_a t + \arctan[\tan(\omega_d t) \cdot (A_1 - A_2)/(A_1 + A_2)])$$

where $\omega_d = (\omega_1 - \omega_2)/2$ and $\omega_a = (\omega_1 + \omega_2)/2$

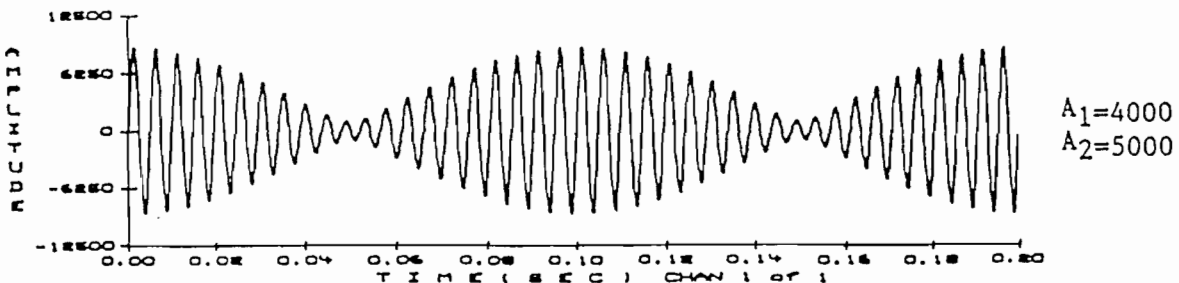
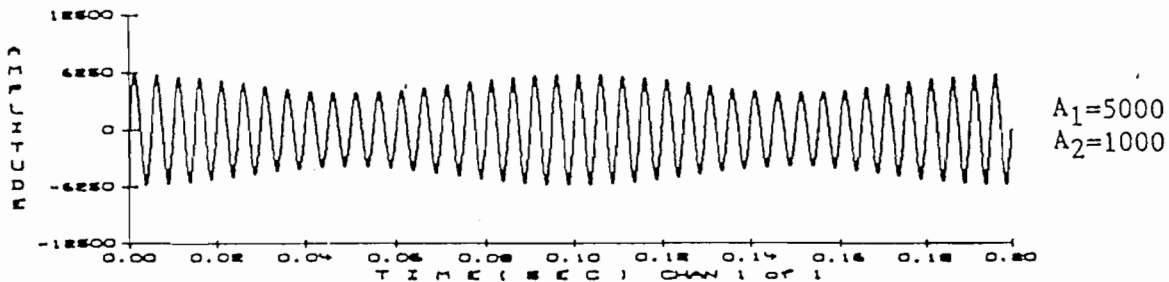
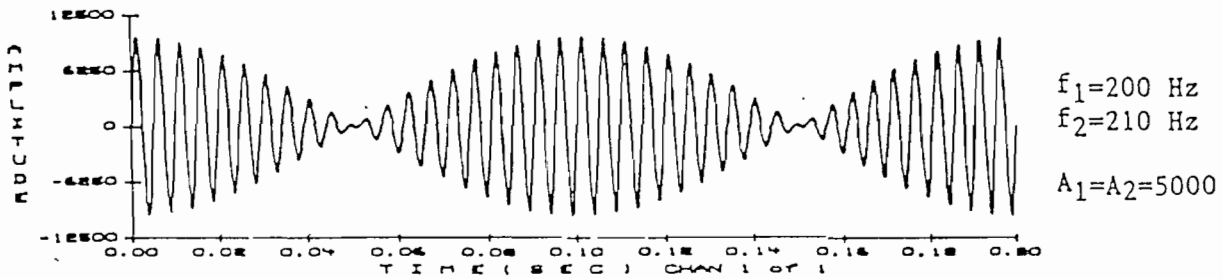


Figure 2a:

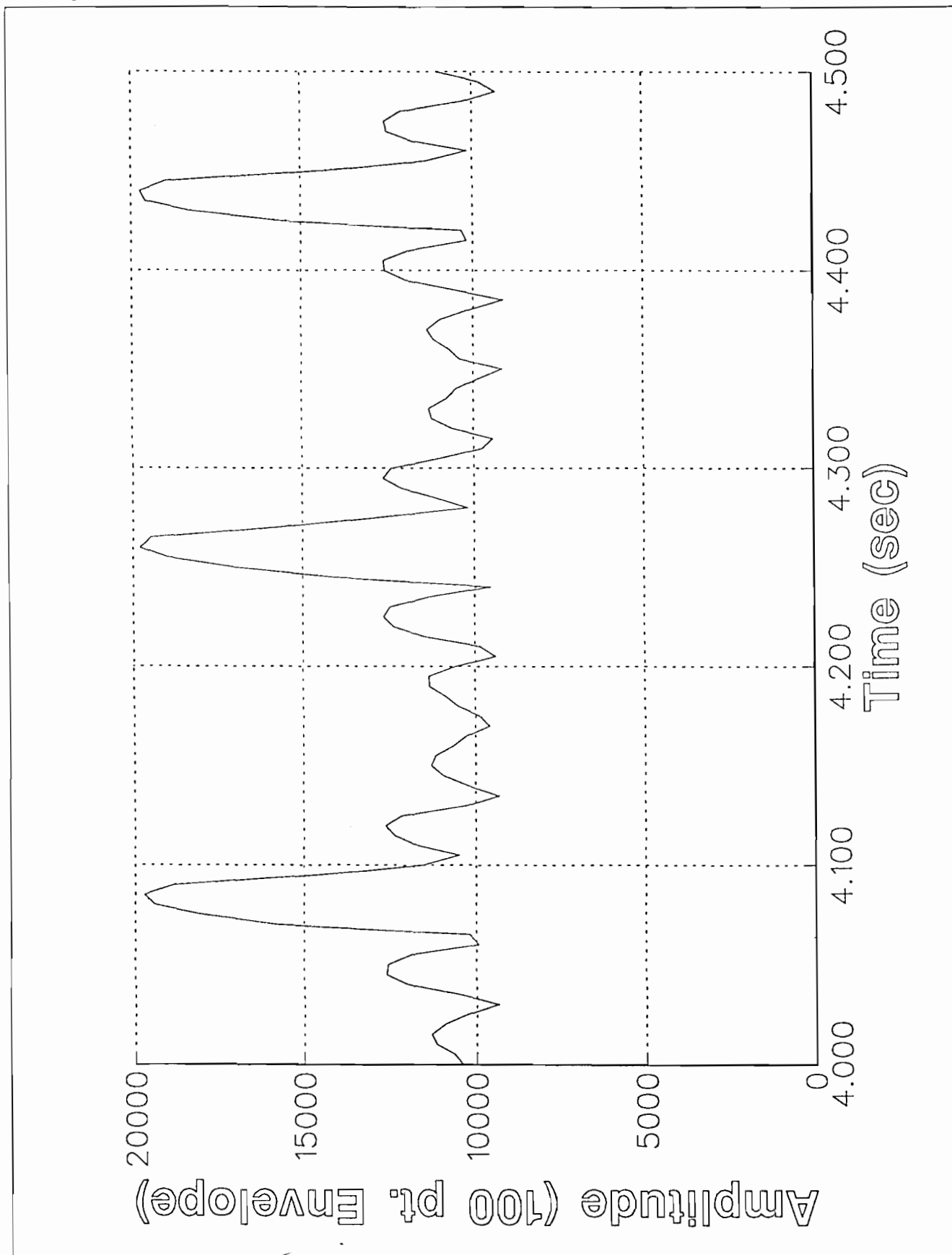


Figure 2b:

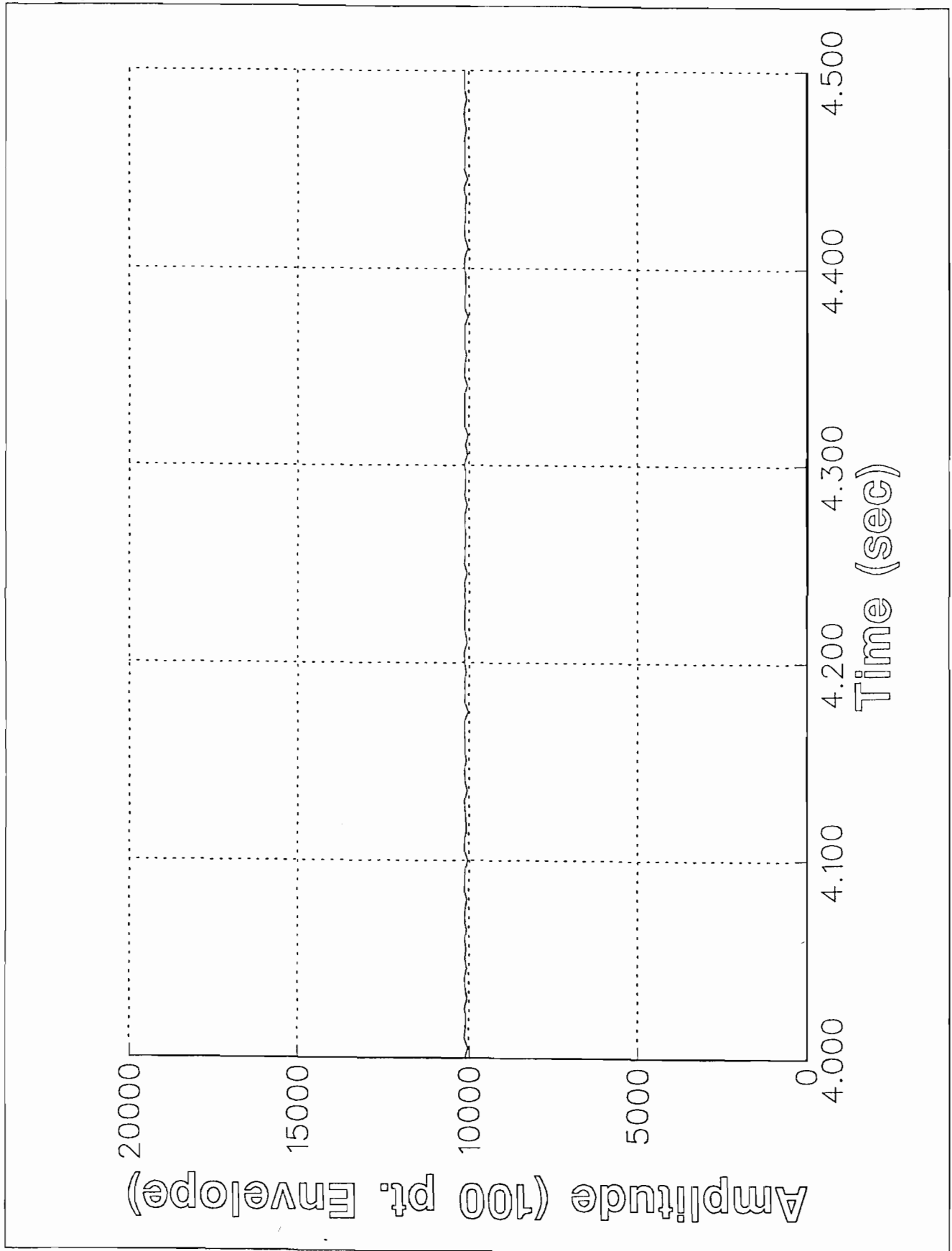


Figure 2c: MQ Analysis for processed mix with beat reduction, example 1.

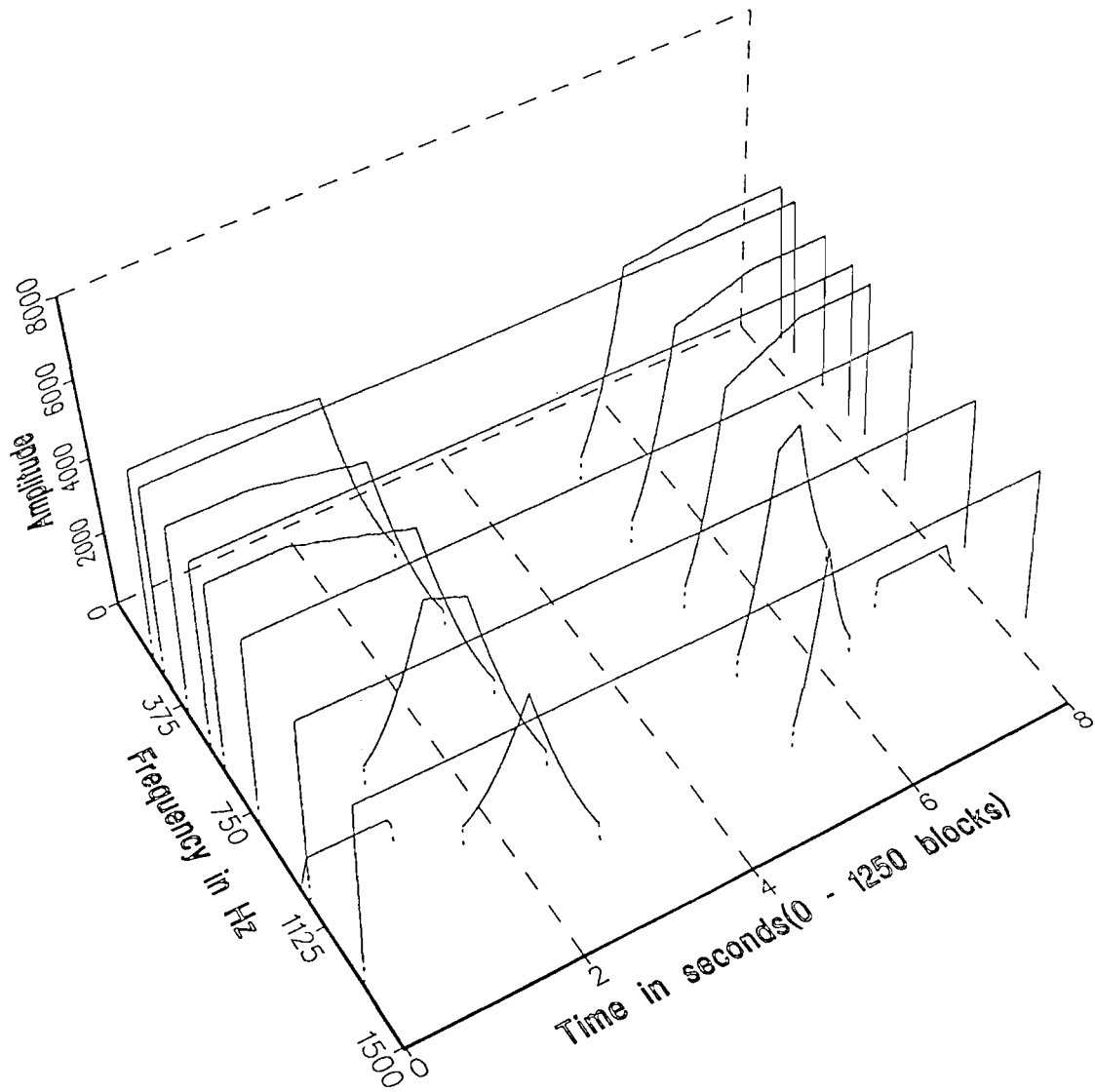


Figure 3a: Exclusion filter for dominant input signal, example 2.

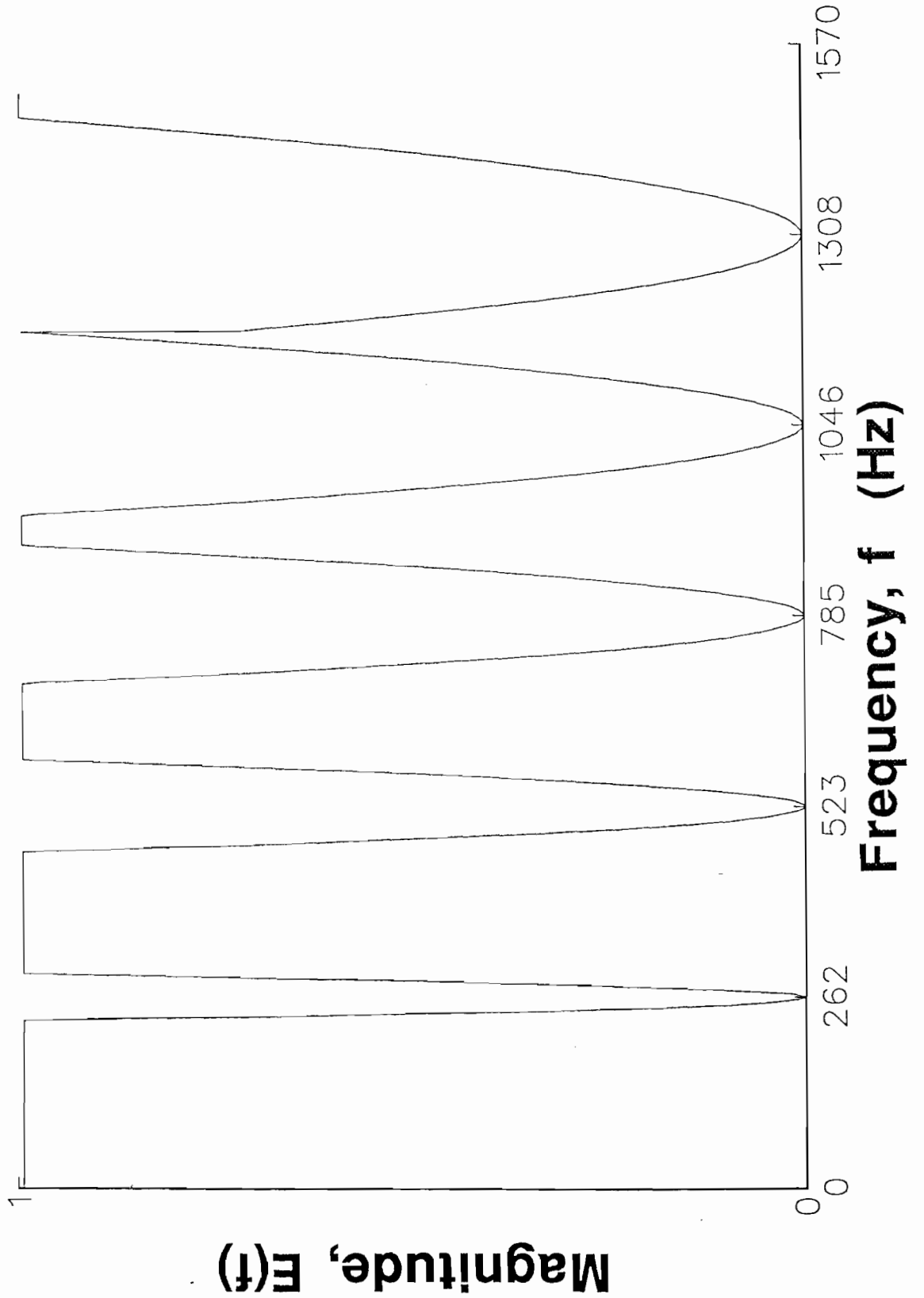


Figure 3b: Processed spectrum for interval C4:G4, example 2.

Component Frequency, f [Hz]	Fundamental Frequency and Partial #	Exclusion Filter Gain, E(f)
261.6	C4, 1	--
392.0	G4, 1	1.0
523.3	C4, 2	--
784.0	G4, 2	0.0001
784.9	C4, 3	--
1046.5	C4, 4	--
1176.0	G4, 3	0.71
1308.1	C4, 5	--
1568.0	G4, 4	1.0
1960.0	G4, 5	1.0

Figure 3c: MQ Analysis for processed mix with beat reduction, example 2.

