

UNIVERSITY OF WESTMINSTER



WestminsterResearch

<http://www.wmin.ac.uk/westminsterresearch>

Tonality index of sigma-delta modulators: a psychoacoustics model based approach.

Jaswinder Lota
Mohammed Al-Janabi
Izzet Kale

School of Informatics

Copyright © [2007] IEEE. Reprinted from the Proceedings of the IEEE International Symposium on Circuits and Systems 2007, (ISCAS 2007). IEEE, Los Alamitos, USA, pp. 729-732. ISBN 1424409217.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Westminster's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners. Users are permitted to download and/or print one copy for non-commercial private study or research. Further distribution and any use of material from within this archive for profit-making enterprises or for commercial gain is strictly forbidden.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of the University of Westminster Eprints (<http://www.wmin.ac.uk/westminsterresearch>).

In case of abuse or copyright appearing without permission e-mail wattsn@wmin.ac.uk.

Tonality Index of Sigma-Delta Modulators : A Psychoacoustics Model Based Approach

Jaswinder Lota ^{†*}, MIEEE, Mohammed Al-Janabi^{*}, MIEEE, Izzet Kale^{*+}, MIEEE

[†]Sepura Ltd, Technology Group, Radio House, St. Andrew's Road, Cambridge CB4 1GR, United Kingdom

^{*}Applied DSP and VLSI Research Group, Department of Electronic Systems, University of Westminster, London, United Kingdom

⁺Applied DSP and VLSI Research Centre, Eastern Mediterranean University, Gazimagusa, N.Cyprus

jasi@ieee.org, {M.Al-Janabi,kalei}@westminster.ac.uk

Abstract – This paper details the development of an enhanced accuracy tonality index for Δ - Σ modulators that use Psychoacoustic techniques. This tonality measure will be especially relevant for quantifying Δ - Σ modulators for performance of audio applications where the output spectra contain low amplitude tones. This tonality index will be very useful for the design and evaluation of Δ - Σ modulators circumventing the need to build a modulator prototype and undertake detailed listening tests.

INTRODUCTION

One of the limitations of Δ - Σ modulators is the existence of tones in their output spectra. This is due to the non-linearity of the 1-bit quantizer which invariably produces non-white quantization noise. The periodic nature of this quantization noise can lead to the formation of tones which are detrimental to the performance of Δ - Σ modulators, especially in audio applications. Simulations and practical experience have shown that the quantization noise present in the output of Δ - Σ modulators tends to contain low-level discrete tones, which are detectable by the human ear [1]. It is shown in [2] that the characterisation of quantization noise by conventional spectral techniques is insufficient for tonal analysis. In [3], an expression for the Tonality Index (T) has been derived based on the power of the output signal spectrum. Listening tests have shown that signals with T less than 0.01 sound like noise, whereas those with T greater than 0.1 contain perceptible tones. The gap between 0.01 and 0.1 is a grey area due to the lack of a connection between the definition of T in [3] and Psychoacoustics [4]-[7]. The aim of this paper is to quantify the audibility of the modulator output signal in this grey area using psychoacoustic principles. The results will assist with the design and performance evaluation of Δ - Σ modulators especially for audio applications. In section II, the expression for T is explained. Section III gives an overview of an algorithm based on psychoacoustic principles used in MPEG/audio coding [9]-[10] to quantify whether a signal is audible or not. Simulation results are given in section IV, followed by conclusions in section V.

TONALITY INDEX

The tonality content in the magnitude spectra of Δ - Σ modulators can be evaluated using T . This is defined as the ratio of the power in the dominant FFT bins to the total power of the input signal. The expression in (1) gives an

adequate approximation of the tonality index for white noise:

$$\tau_{bias} \approx \frac{2m}{N} \left[\ln \left(\frac{N}{2m} \right) + 1 \right] \quad (1)$$

where N is the FFT length and m is the number of dominant bins [3]. Using a large value of $2m/N$ would result in a greater bias and a larger uncertainty in T . In order to account for this skewing effect, the estimated bias is subtracted and the resulting index renormalised. The resulting tonality index is given by:

$$T = \frac{\tau_o - \tau_{bias}}{1 - \tau_{bias}} \quad (2)$$

where τ_o is the raw tonality index. The expression in (2) gives a reasonable evaluation of tonality but is rather crude, as it has no relation to the human auditory system. The human auditory system is non-linear and has its peculiarities. The minimum signal threshold for audibility changes, as the signal frequency increases from 1 to 20 kHz. In addition, the signal at a particular frequency may have masking effects on the nearby frequencies thereby affecting audibility [4].

PSYCHOACOUSTIC PRINCIPLES

Psychoacoustics has made significant progress in quantifying human auditory perception. Using psychoacoustics, a signal can be classified as audible or inaudible. This categorization can be then used for the coding of audio signals also known as perceptual coding. The psychoacoustic model delivers masking thresholds, which quantify the maximum amount of distortion that can be injected during quantization without introducing audibly perceivable distortion in the reconstructed signal. Most audio coders achieve compression by exploiting the information that is not detectable by the human ear. This irrelevant information is identified when using psychoacoustic principles such as absolute hearing thresholds, critical band frequency analysis, simultaneous masking and masking spread along the basilar membrane [8]. The psychoacoustic model analyzes the audio signal and computes the amount of tonality or noise masking available as a function of frequency. The masking ability of a given signal component depends on its frequency and loudness. With the noise and signal masking values, the Global Masking Threshold (GMT) is computed. The signal

excess level above the GMT determines the audibility of the signal.

The two psychoacoustic models used in MPEG/audio coding standards are Psychoacoustics Model-I & Psychoacoustics Model-II [9]-[10]. Both models work for any of the MPEG layers of compression. Model-I is used in this paper for tonal analysis. An excellent description of Model-I is given in [11] and the basic steps of its algorithm are detailed below.

STEP I: Spectral Analysis

The incoming signal samples $y(n)$ are normalised as per the FFT length N which references the power spectrum to a 0 dB maximum. The normalised input $x(n)$ is then segmented into 12 ms frames (512 samples) using a $1/16^{\text{th}}$ overlapped Hanning window $w(n)$. The Power Spectral Density (PSD) $P(k)$ of the signal is then estimated using a 512 point FFT:

$$P(k) = 90 + 10 \log_{10} \left| \sum_{n=0}^{N-1} w(n)x(n)e^{-j\frac{2\pi kn}{N}} \right|^2 \quad 0 \leq k \leq N/2 \quad (3)$$

where 90 dB is the power normalization term. After normalization the signal power level units become known as dB Sound Pressure Level (SPL).

STEP II: Identification of Tonal and Noise Maskers

After the PSD estimation, the tonal and noise masking components are identified. Local maxima in the PSD samples which exceed the neighbouring samples by 7 dB are classified as tonal. The tonal set can be defined as:

$$\chi_T = \left\{ P(k) = \begin{cases} P(k) > P(k \pm 1), \\ P(k) > P(k \pm \tau) + 7 \text{ dB} \end{cases} \right\} \quad (4)$$

where

$$\tau \in \begin{cases} 2, 2 < k < 63 \\ [2, 3], 63 < k < 127 \\ [2, 6], 127 < k < 256 \end{cases} \quad (5)$$

Tonal maskers $T_M(k)$ are computed from the spectral peaks listed in χ_T as follows:

$$T_M(k) = 10 \log_{10} \sum_{j=-1}^1 10^{0.1P(k+j)} \text{ dB} \quad (6)$$

A single noise masker for each critical band, $N_M(k')$ is computed from the remaining spectral lines not within the $\pm\tau$ neighbourhood of a tonal masker using the expression:

$$N_M(k') = 10 \log_{10} \sum_j 10^{0.1P(j)} \text{ dB} \quad (7)$$

$$\forall P(j) \notin \{T_M(k, k \pm 1, k \pm \tau)\} \quad (8)$$

where k' is defined to be the geometric mean spectral line of the critical band,

$$k' = \left(\prod_{j=l}^u j \right)^{1/(l-u+1)} \quad (9)$$

where l and u are the lower and upper spectral line boundaries of the critical band respectively.

STEP III: Decimation & Reorganisation of Maskers

Here any maskers below the absolute threshold are discarded. Only those maskers are retained that satisfy the following relationship:

$$T_M(k), N_M(k) \geq N_q(k) \quad (10)$$

where $N_q(k)$ is the SPL of the listening threshold. Next a sliding 0.5 bark (defined in (15)) wide window is used to replace any pair of maskers occurring within a distance of 0.5 bark with the stronger one. Subsequently, the masker frequency bins are reorganized according to the sub-sampling scheme.

$$T_M(i), N_M(i) = T_M(k), N_M(k) \quad (11)$$

$$T_M(k), N_M(k) = 0 \quad (12)$$

where

$$i = \begin{cases} k, 1 \leq k \leq 48 \\ k + (k \bmod 2), 49 \leq k \leq 96 \\ k + 3 - ((k - 1) \bmod 4), 97 \leq k \leq 232 \end{cases} \quad (13)$$

STEP IV: Calculation of Individual Masking Thresholds

Individual tone and masking thresholds are computed next. Each individual threshold represents a masking contribution at frequency bin i due to the tone or noise masker located at bin j (reorganized in the step above). Tonal masker thresholds $T_T(i, j)$ are given by:

$$T_T(i, j) = T_M(j) - 0.275z(j) + SF(i, j) - 6.025 \text{ dB} \quad (14)$$

where $T_M(j)$ denotes the SPL of the tonal masker in frequency bin j , $z(j)$ denotes the Bark frequency of bin j which for a frequency of f in Hertz given by:

$$Z(f) = 13 \arctan(0.00076f) + 35 \arctan\left[\left(\frac{f}{7500}\right)^2\right] \text{ (Bark)} \quad (15)$$

the spread of the masking from the masker bin j to maskee bin i , $SF(i, j)$, is modelled by the expression:

$$SF(i, j) = \begin{cases} 17\tau - 0.4T_M(j) + 11, -3 \leq \tau < -1 \\ (0.4T_M(j) + 6)\tau, -1 \leq \tau < 0 \\ -17\tau, 0 \leq \tau < 1 \\ (0.15T_M(j) - 17)\tau - 0.15T_M(j), 1 \leq \tau < 8 \end{cases} \quad (16) \quad (dB)$$

which is a piecewise linear function of masker level, $T(j)$ and bark maskee-masker separation, $\Delta_z = z(i) - z(j)$. Individual noise masker thresholds $T_N(i, j)$ are given by:

$$T_N(i, j) = N_M(j) - 0.175z(j) + SF(i, j) - 2.025dB \quad (17)$$

where $N_M(j)$ denotes the SPL of the noise masker in the frequency bin j , and $SF(i, j)$ is obtained by replacing $T_M(j)$ by $N_M(j)$ everywhere in (16).

STEP V: Calculation of Global Masking Thresholds

The individual masking thresholds are combined to estimate a global masking threshold for each frequency bin in the subset given in (13). The global masking threshold T_G is obtained by computing the sum

$$T_G = 10 \log_{10} \left(10^{0.1N_q(i)} + \sum_{l=1}^L 10^{0.1T_l(i,l)} + \sum_{m=1}^M 10^{0.1T_N(i,m)} \right) \quad (18)$$

where $N_q(i)$ is the absolute hearing threshold for the frequency bin i , $T_l(i, l)$ and $T_N(i, m)$ are the individual masking thresholds in the previous step, l and m are the number of tonal and noise maskers respectively. The global threshold for each frequency bin represents a signal dependent, power additive modification of the absolute threshold due to the basilar spread of all tonal and noise maskers in the signal power spectrum.

SIMULATIONS & RESULTS

Simulations were undertaken in MATLAB for a second-order discrete-time Δ - Σ modulator for audio applications with a DC input power of -60 dB as shown in Figure 1. The sampling frequency was 1.344 MHz for an oversampling ratio (OSR) of 64. The coefficient values a_1 , a_2 , a_3 and a_4 were 0.2, 0.2, 0.5 and 0.25 respectively. The output was filtered by a low-pass FIR filter with a cut-off

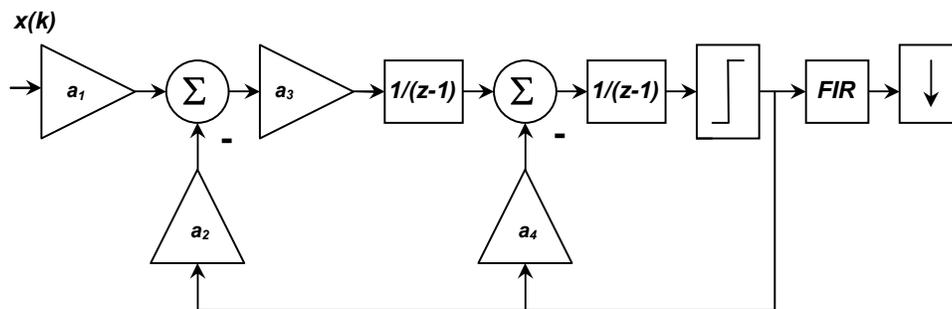


Figure 1. 2nd-order Discrete-Time Δ - Σ Modulator for audio applications

frequency of 21 kHz and decimated by a factor of 64. The filter was realized as a Tapped delay line having the order 348. The input signal was additively contaminated with Additive White Gaussian Noise (AWGN) with increasing values of variance in order to decrease the level of tones. This effectively reduced the value of T from 0.1 to 0.01. Listening tests for the simulations were undertaken by the MATLAB Signal Processing Tool (sptool), Graphical User Interface (GUI).

Steps I-V of the algorithm above were implemented with a 512-point FFT. The FFT samples $P(k)$ along with $T_M(k)$, $N_M(k)$ and $N_q(k)$ are shown in Figure 2. From (1) and (2), T was found to be 0.406 for $N=16384$ and $m=9$. It is seen that although tones VII-IX contribute to the numerical value of T , they are below $N_q(k)$ and therefore would be inaudible. The tones VII-IX therefore do not contribute to the tonality of the spectrum, although they theoretically contribute towards the numerical computation of T in (1) and (2). In effect the spectrum is perceived as tonal due to the contributions of tones I-VI ($m=6$) only, but has an incorrectly computed value of T for contributions from all the tones I-IX ($m=9$). The existing definition of T is unable to highlight this fact. If the spectrum had consisted only of the tones VII-IX ($m=3$), the computed value of T would have been 0.367, which is regarded as tonal. However to a listener the spectrum would not be perceived as tonal as it would be inaudible.

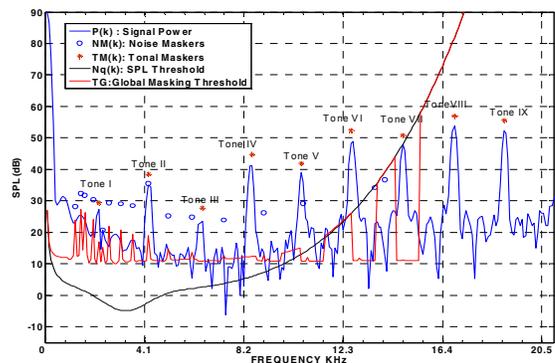


Figure 2. FFT signal samples for 9 dominant tones

For the signal input power of -28 dB the spectrum for $T = 0.002$ and $m = 2$ is plotted in Figure 3. Tones I and II are in effect not classified as tones as they do not satisfy the conditions (4) and (5) of Step II. The signal is therefore perceived as absolute noise by a listener.

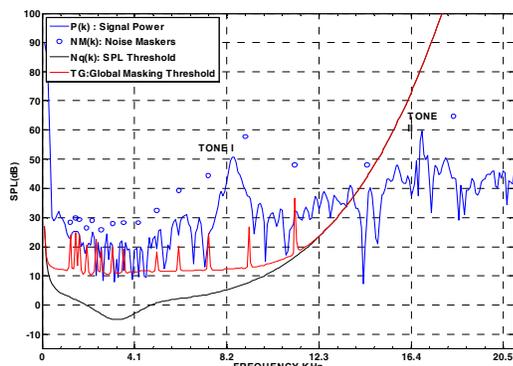


Figure 3. FFT signal samples for 2 dominant tones for $T=0.002$

For $T = 0.08$ and $m = 2$, the spectrum is shown in Figure 4. Tone I is classified as a tone while Tone II as noise. The spectrum would be perceived as tonal due to Tone I. The SPL excess above the T_G for the tone is 30.43 dB. Even if Tone II were to be classified as a tone it would not contribute to the tonality as it is below the $N_q(k)$.

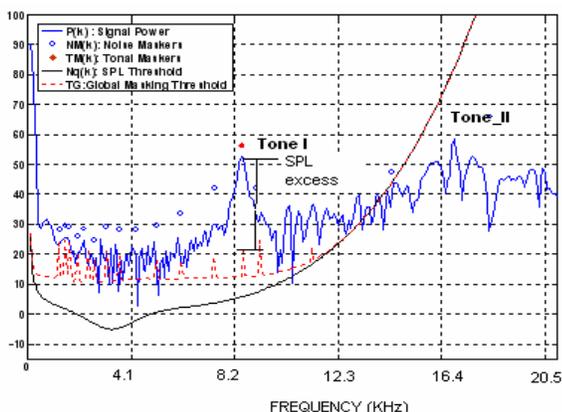


Figure 4. FFT signal samples for 2 dominant tones for $T = 0.08$

Figure 5 shows the plot of the SPL excess level against the value of T in the region 0.01 to 0.1 for $m = 2$. For $T < 0.04$ the spectrum is audible as noise as neither of the tones I and II are classified as tonal. For $T > 0.04$, tone I is classified as tonal and its SPL excess level is found to be 30 dB. The SPL excess level remains constant up to $T = 0.1$. In the region $T > 0.04$, the spectrum would be tonal provided the tone power is greater than the quiet listening threshold $N_q(k)$. This threshold is about 40 dB at 13 kHz and increases steeply thereafter. The occurrence of such spurious tones (with power greater > 40 dB) is unlikely. Therefore the probability of the spectrum being tonal at frequencies greater than 13 kHz is low even for $T > 0.04$.

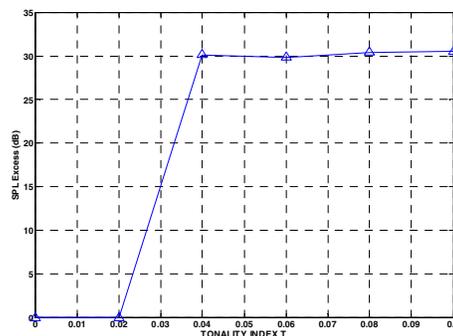


Figure 5. SPL excess for $0.01 < T < 0.1$

CONCLUSIONS

An adequate tonal measure of a signal with T in the range of 0.01-0.1 using psychoacoustics has been established. Listening tests were undertaken with the MATLAB spool GUI to validate the results. It is seen that signals with values of $T < 0.04$ can be regarded as noise. The signals with $T > 0.04$ can be regarded as tonal depending on their signal power and frequency. By definition since T is dependent only on the signal power and not frequency, it may incorrectly quantify the signals as tone and is therefore unable to resolve tonality of signals in the region 0.01-0.1. As no such metric exists for the analysis of such low-level tonal content signals, the measure will provide invaluable guidance with respect to listening or audio performance without the need to build a prototype. The effects of non-idealities on tonality of discrete-time $\Delta\Sigma$ modulators will be reported in a future publication.

REFERENCES

- [1] M. Motamed, A. Zakhor, S. Sanders, "Tones, Saturation & SNR in Double-Loop Delta-Sigma Modulators," in *Proc. IEEE ISCAS 93*, pp. 1345-1348, Chicago, IL USA, May 1993.
- [2] R. Schreier, "Destabilising Limit Cycles in Delta-Sigma Modulators with Chaos," in *Proc. IEEE ISCAS 93*, pp. 1369-1372, Chicago, IL USA, May 1993.
- [3] R. Schreier, "On the Use of Chaos to reduce Idle-Channel Tones in Delta-Sigma Modulators," *IEEE Trans. Circuits & Syst-I*, vol.41, No.8, August 1994.
- [4] H. Fletcher, "Auditory Patterns", *Rev. Mod. Phys.*, pp. 47-65, January 1940.
- [5] D.D. Greenwood, "Critical Bandwidth and the Frequency Coordinates of Basilar Membrane," *J.Acoust. Soc. Am.*, pp. 1344-1356, October 1961.
- [6] M. Schroeder, "Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear," *J. Acoust. Am.*, pp. 1647-1652, December 1979.
- [7] E. Zwicker, U. Zwicker, "Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System," *J. Audio Eng. Soc.*, pp. 115-126, March 1991.
- [8] L. Pikula, "Objective Perceptual Evaluation of the Quality of Audio Signals", www.mms.ecs.soton.ac.uk/papers/lp1403.rtf.
- [9] Davis Pan, "A Tutorial on MPEG/Audio Compression", *IEEE Multimedia 1995*.
- [10] Peter Noll, "Wideband Speech and Audio Coding", *IEEE Communications Magazine*, November 1993.
- [11] T. Painter, A. Spanias, "A Review of Algorithms for Perceptual Coding of Digital Audio Signals", *Int. Conf. DSP, DSP-97*, Hellas, Greece, July 1997.