FIFTH INTERNATIONAL CONGRESS ON SOUND AND VIBRATION

DECEMBER 15-18, 1997
ADELAIDE, SOUTH AUSTRALIA

A PROOF OF THE VARIANCE FORMULA FOR THE TOTAL CROSSING TIME OF A CONTINUOUS RANDOM SOUND SIGNAL WITH RESPECT TO A FIXED LEVEL

A. Heiss

**Bavarian Ministry of Environmental Protection,
Rosenkavalierplatz 2, D-81925 Munich, Germany**

ABSTRACT

For a quality controlled application of the percentile sound evaluation index L_x , its accuracy, limited by the stochastic level fluctuations, should be quantified. This is achievable by determining first the variance of the partition of the signal amplitude with respect to a fixed level, the expectation value of L_x . Then the percentile's spread itself is accessible through the cumulative level distribution.
As is still to be done, this paper presents a comprehensible proof of the partition variance formula, which is of fundamental importance for the evaluation of the L_x - confidence limits. Given a number of crossings within a given measurement time interval the probability that a definite total of the single continuous overshoots, i. e. "crossing up" time intervals occurs, depends as on the stochastic system's crossings *up* as on its crossings *down*, and further on the probability density function (p. d. f.) of the crossing number. Above an easily practicable minimum crossing number the p. d. f. of the total crossing time, and so of the partition itself, can be presented explicitely and straightforward in microstatistic terms applying the Central Limit Theorem of Statistics. Then step by step integration of the variance definition equation leads to the already known and applied final result.

1. INTRODUCTION

In practice of environmental noise control noise levels are encountered which, as a rule, fluctuate randomly - often over several decades of decibels even during relatively short time intervals. It is evident that for the exhaustive description of those fluctuations x%-quantiles, i. e. the percentile levels L_x , are the appropriate kind of measurement index. For this reason already since many years the sound percentiles are part of the standard measurement technique.

Due to the stochastics being at work the measured L_x value is of restricted accuracy, like any other kind of measurement index. Its nonzero spread can be visualized by a manyfold repetition of the measurement under nonvarying measurement conditions.
In a former paper of the author is demonstrated how the latent spread of a single measured L_x value can be described by the underlying statistical structural parameters of the signal and of the length of the measurement time interval [2]. The access to the percentile spread is possible by taking advantage of the variance of the partition imposed by the L_x expectation value on the stochastic sound signal's instantaneous values. The transition to the finally interesting bracket confidence interval calculated according to a given confidence level [1] is accomplished by use of the cumulative distribution of the sound level [2][3].
The purpose of this present paper is to derive in comprehensive detail the partition variance formula which for the first time was used in the short basic paper [2].

2. THE STOCHASTIC SYSTEM

Let us use for simplicity instead of x % the parameter $q \in \mathbf{R}^+$, $0 \le q \le 1$. It denotes the fraction of time during which a given value of a magnitude like the sound pressure level, here denoted by L_q , is *exceeded* by the instantaneous amplitudes. The mutual assignment of the variables q and L_q is established by the definition equation

$$q(L_q) = \frac{1}{T}\sum_{i=1}^{n} w_i := \frac{W}{T} . \qquad (1)$$

The further notations used in this expression and illustrated by Fig. 1 are: T: Measurement time interval; n: In T observed number of the time intervals w_i , the "crossing up" intervals created by the immediately successive crossing up and crossing down of the time dependent sound pressure level relative to L_q . L_q is here presupposed as an expectation value. W denotes the sum of the crossing up intervals. Its number should be $n \gtrapprox 7$. The reason for this will be given below. Since w_i , n and consequently W are stochastic variables each of them evidently obeys an own probability density function (p.d.f.) [1]. The most simple physical model which can create the quantities of kind w_i at a fixed level is a sequence of single transient sound events fully separated in time like the single passing by of vehicles which occurs in case of traffic with very low intensity.

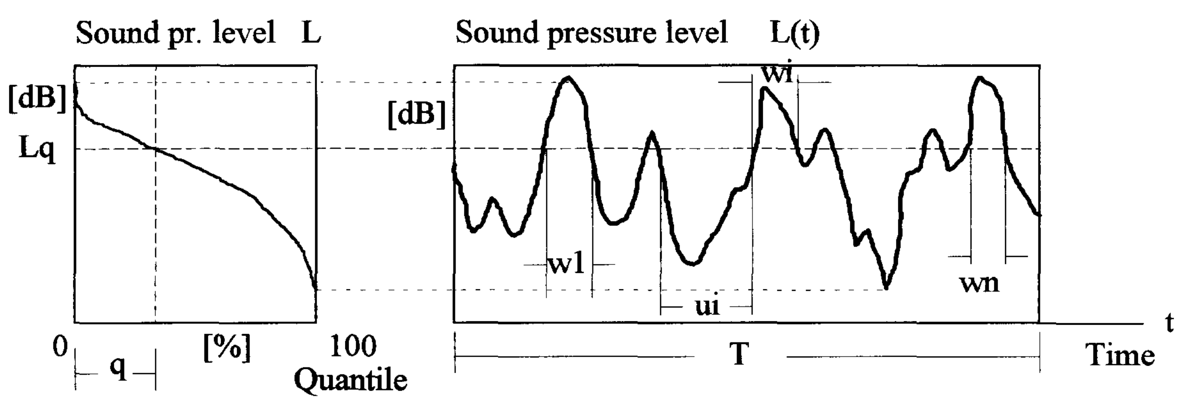

Fig. 1: Definition of the parameters and stochastic variables used in eq. (1).
For the derivation of the partition variance formula the time dependent level value L_q is to be considered for the moment as a representative fixed value, i. e. as its expectation. (Schematic presentation.)

3. STATE DENSITY EQUATIONS

3.1 Distribution of the total crossing time

If the variance of W, the total of crossing up intervals is accessible, then according to eq. (1) also the variance of the partition q is known. Then using Var q the variance of the interesting quantile (percentile) level L_q is accessible by an unambiguous procedure from the cumulative distribution function of the time dependent sound pressure level signal. This is already demonstrated in full detail in [3].
The variance of W dependent on the system's parameters is by its basic definition [1]

$$Var\ W \quad = \quad \overline{(W-\overline{W})^2} \quad = \quad \overline{W^2} - \overline{W}^2 . \tag{2}$$

As usual the bar in eq. (2) indicates the expectation value. It is per definition the mean taken over the distribution of all possible outcomes of the variable W ([1], Ch. 1.1). Var W is determined over the sample space $S = \{W : 0 \le W \le T\}$.
For further evaluation let us consider first the case of a given value for an integer $n \ge 2$. Then the p.d.f. $\varphi(W,n)$, where n has the function of a set parameter, is completely determined by the n-1-fold convolution ([4], Ch. 12.2)

$$\varphi(W,n) = \int\limits_{w_{n-1}=0}^{W} \int\limits_{w_{n-2}=0}^{W-w_{n-1}} \cdots \int\limits_{w_1=0}^{W-w_2-\ldots-w_{n-1}} g(W-w_1-\cdots-w_{n-1}) g(w_1) \cdots$$
$$\cdots g(w_{n-2}) g(w_{n-1})\, dw_1 \cdots dw_{n-2} dw_{n-1} \tag{3}$$

where $g(w)$ denotes the p.d.f. (normalized to unity) of the single crossing up intervals w_i preassumed to be stochastically independent..
Distributions of w_i and u_i extracted from everyday field measurements of environmental noise show that these probability densities can approximated quite well by a function of the type

$$g_n(w) = x^{n-1}\, e^{-x} \,/\, \Gamma(n) , \tag{4}$$

where typically $1/2 <\approx n <\approx 3/2$. Due to the inherent system property

$$W + U \equiv T = const. \tag{5}$$

W and the total U of the crossing down intervals u_i (see Fig. 1) are strongly connected by the measurement time interval. They have to share the same available time reservoir T among themselves. Also the variable from the type u is assumed to be stochastically independent. The most simple physical models that can create the quantities u_i are the time distances between the separate passing by events of aircrafts or of vehicles if there is traffic with very low intensity.
As the u_i and w_i always occur in an immediate alternation, they generate "stochastic periods", for example in Fig. 1 from the beginning of u_i to the end of w_i . Hence to the

variable of type u the same number n can be attributed as to the variable of type w . According to these system properties, primarily the kind of conditional probability which is acting here, the final p.d.f. for W now can be expressed. For this purpose we also need the p.d.f. of the crossing down total U denoted here by $\psi(U,n)$. This p.d.f. can be calculated completely analogous to eq. (3) by replacement of w and g(w) by u and its p.d.f. As the basic p.d.f. g(w) is in general a continuous function by eq. (3) also the p.d.f. $\varphi(W,n)$ is continuous in W . Thus if $\varphi(W,n)$ is also clearly dependent on n it can be expressed as a continuous function of n . This corresponds with reality where usually a noninteger number of stochastic periods is cut out by the measurement interval T . Hence the final state density of W , i. e. its p.d.f. denoted by $\phi(W,n,T)$, can be expressed as

$$\begin{aligned} d^2P(W,n,T) \;&:=\; P\{W \le W' \le W + dW; n \le n' \le n + dn \mid W + U = T\} \\ &=\; a(n,T)\cdot\varphi(W,n)\,\psi(T-W,n)\,dW\cdot f(n,T)\,dn \qquad (6) \\ &\xrightarrow{def}\; \phi(W,n,T)\,dW\,dn\,. \end{aligned}$$

In eq. (6) a(n,T) denotes a factor which normalizes $\varphi(W,n)\cdot\psi(T-W,n)$ to unity over the parameter space $S_W = \{W : 0 \le W \le T\}$. The p.d.f. of the crossing number $n \in \mathbf{R}^+$ is denoted by f(n,T).

3.2 Explicite representation by expectation values using the Central Limit Theorem

In favor of a comfortable presentation and evaluation of the eqs. (3), (6) and (2) in terms of easy accessible observables the basic expectation values $\bar{u}$, $\bar{w}$,

$$\bar{n} \;:=\; T/(\bar{u}+\bar{w}) \qquad (7)$$

and also the variances σ_u of u_i and σ_w of w_i are used.
In eq. (1) we have a linear sum of the crossing up vector components w_i . If there is a sufficient high number of crossings in T , the sum in eq. (1) which is distributed in general as expressed by eq. (3), becomes approximately a normal distribution, what is the well known content of the Central Limit Theorem (C.L.T.) ([1], ch. 5.3). The mean value parameter of this p.d.f. is

$$\overline{W} \;=\; \bar{n}\cdot\bar{w} \qquad (8)\,.$$

The variance parameter is

$$\sigma_W^2 \;=\; \bar{n}\cdot\sigma_w^2\,. \qquad (9)$$

The normal distribution

$$N_x(\mu,\sigma^2) \;\equiv\; N(x;\mu,\sigma^2) \;=\; \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \qquad (10)$$

([1], ch. 3.4) is specified here by the index x representing symbolically the variables W or n . As in [1] is cited generally, the approximation by the C.L.T. will be good if n is greater than 25 or 30. It can easily be shown by computer aid that if the distribution is of the single sample type eq. (4) , already a value of n as small as 7 yields an adequate approximation: In this case the deviation of the confidence limits [1][2][3] from their values which result from the exact p.d.f. is less than 10 % and thus tolerable. Hence the p.d.f. $\varphi(W,n)$, eq. (3) can be denoted as

$$\varphi(W,n) = N(n\bar{w}, n\sigma_w^2) \tag{11}$$

and corresponding $\psi(U,n) = \psi(T-W,n)$ by

$$\psi(T-W,n) = N(T-n\bar{u}, n\sigma_u^2) \ . \tag{12}$$

From these two eqs., after some algebraic rearrangement, follows

$$\varphi(W,n)\ \psi(T-W,n) = \nu_P \cdot N_W\left[\frac{n\bar{w}\,\sigma_u^2 + (T-n\bar{u})\sigma_w^2}{\sigma_u^2+\sigma_w^2},\ n\frac{\sigma_u^2\sigma_w^2}{\sigma_u^2+\sigma_w^2}\right] \cdot N_n\left[n, \sigma_n^2\right] \tag{13}$$

where

$$\nu_P := 1/(\bar{u}+\bar{w}) \tag{14}$$

is the mean frequency of crossing up and crossing down intervals respectively. The number n is still a freely varying parameter.

It is remarkable that in consequence of the constraint eq. (5) the variance of W is primarily determined by the smaller of the two variances σ_u and σ_w. For the crossing number the expectation $\bar{n}$ is determined according to eq. (7) and its variance by

$$\sigma_n^2 = n\frac{\sigma_u^2+\sigma_w^2}{(\bar{u}+\bar{w})^2} \tag{15}$$

which is composed additively by two stochastically independent terms. The phenomenon "number variance" has already been mentioned and shortly discussed in [6].

In favor of a relatively simple performance to present the right hand of eq. (6) in terms of the basic parameters T , n resp. $\bar{n}$, $\bar{u}$, $\bar{w}$, σ_u and σ_W it is convenient - and also justified by the measurement conditions in the daily practice - to presuppose that the measurement time interval T covers N_W [...] almost completely. For control that this condition is met, an adequate criterion for the required minimum crossing number is to be established. With respect to the nearer, i. e. the crucial edge of the sample space to cut off the p.d.f. this criterion is

$$\mu \geq z_{1-\alpha/2} \cdot \sigma \, . \tag{16}$$

In eq. (16) z denotes the quantile of the N(0,1) distribution which results from a partition 1-α/2. According to this α indicates the maximum possible deviation of an integration over the distibutions in eq. (13) in comparison with an unbounded integration which yields unity.

If the corresponding parameter included in eq. (13) are inserted in eq. (16), in the worst case of W for the upper edge of the sound pressure level distribution, i. e. if $\bar{w} << \bar{u}$ and under the assumption that the variance coefficients of the stochastic variables u_i and w_i i. e.

$v_u := \sigma_u / \bar{u}$ and $v_w := s_w / \bar{w}$ do not exceed the order of magnitude 1 considerable, then easily can be shown that the general condition for the minimum number required is simply

$$n \geq z^2_{1-\alpha/2} \, . \tag{17}$$

For an approximation of the normal distribution for example to at least 99 % (i. e. $\alpha \leq 0{,}01$), $z = 2{,}6$ is necessary and n has to be 7 at least. If the condition (17) is also met for n , this parameter can in N_n [...] and σ^2_n be replaced by $\bar{n}$ (see eqs. (13) and (16)). If the condition (17) holds and eq. (13) is inserted into the definition eq. (6) the normalizing of the W-part of ϕ(W,n,T) to unity yields

$$a(T,n) \cdot \bar{v}_P \cdot N_n\left[\bar{n}, \sigma^2_n\right] = 1 \, . \tag{18}$$

With this normalizing condition we get

$$\phi(W,n,T) = N_W[\cdots] \cdot f(n,T) \tag{19}$$

For the calculation of the first and second moments of n to be performed, the p.d.f. f(n,T) must not necessaryly be known explicitely. Only the mean and the variance of n are needed. Despite of this the explicite function f(n,T), at least if condition (17) is met, is of interest. The cumulative probability F(n,T) of the sum of n periods u_i+w_i dependent on the length of time interval T over which it is established, can be expressed by

$$F(n,T) = \int_{t'=0}^{T} \int_{W'=0}^{W} \varphi(W',n) \, \psi(t'-W',n) \, dW' \, dt' \, . \tag{20}$$

If (17) holds then the integration over W yields the normal distribution $N_{t'}\left[n \cdot (\bar{u} + \bar{w}), \ n \cdot (\sigma^2_u + \sigma^2_w)\right]$. From the corresponding cumulative distribution function Φ(T,n) the p.d.f. f(n,T) can be derived by

$$f(n,T) = -\frac{\partial \Phi(T,n)}{\partial n} = -\int_0^T \frac{\partial N_{t'}(t'-n \cdot (\bar{u} + \bar{w}))}{\partial n} dt' = N_n(\bar{n}, \sigma^2_n) \, . \tag{21}$$

Thus the final form of ϕ(W,n,T) is

$$\phi(W,n,T) = N_W\left[\frac{n\bar{w}\,\sigma^2_u + (T - n\bar{u})\sigma^2_w}{\sigma^2_u + \sigma^2_w}, \ n\frac{\sigma^2_u \sigma^2_w}{\sigma^2_u + \sigma^2_w}\right] \cdot N_n\left[\bar{n}, \sigma^2_n\right] . \tag{22}$$

It is here to be emphasized that in eq. (22) N_W , in extension of the strong definition of the normal p.d.f., *still* depends on n .

4. VARIANCE OF THE CROSSING TIME IN TOTAL AND OF THE PARTITION

All components of the p.d.f. $\phi(W,n,T)$, neccessary for the further operations, now are principally known, based on the distributions of the w_i and the u_i variables. Using eq. (22) the first and second moment of W , the crossing up in total principally can be calculated by integration over W and n . These moments are required to determine explicitely Var W according to eq. (2).
The first moment of W is to be calculated by integrations of (22) and using eq. (7) . It is simply

$$\overline{W} = \overline{n}\cdot\overline{w} = \frac{\overline{w}}{\overline{u}+\overline{w}}\cdot T \equiv q\cdot T . \tag{23}$$

To evaluate the second moment the integration over W yields

$$\overline{W^2}(n) = \left(\frac{n(\overline{w}\,\sigma_u^2 - \overline{u}\,\sigma_w^2) + T\sigma_w^2}{\sigma_u^2+\sigma_w^2}\right)^2 + n\,\sigma^{*2} . \tag{24}$$

For abbreviation σ^* is introduced by

$$1/\sigma^{*2} = 1/\sigma_u^2 + 1/\sigma_w^2 . \tag{25}$$

Integration of (22) over n , using the known moment properties of the normal distribution, yields

$$\overline{W^2} = \int_0^\infty \overline{W^2}(n)\cdot N_n(\overline{n},\sigma_n^2)\,dn$$
$$= \frac{1}{(\sigma_u^2+\sigma_w^2)^2}\left[(\overline{n}^2+\sigma_n^2)(\overline{w}\,\sigma_u^2-\overline{u}\,\sigma_w^2)^2 + 2\overline{n}T\sigma_w^2(\overline{w}\,\sigma_u^2-\overline{u}\,\sigma_w^2) + T^2\sigma_w^4\right] + \overline{n}\,\sigma^{*2} \tag{26}$$

Hence according to eqs. (25), (23) and (2) and after an algebraic rearrangement we arrive finally at

$$Var\ W = \overline{n}\left[\left(\frac{\overline{u}}{\overline{u}+\overline{w}}\right)^2\sigma_w^2 + \left(\frac{\overline{w}}{\overline{u}+\overline{w}}\right)^2\sigma_u^2\right] \tag{27}$$

and due to eq. (1) at

$$Var\ q = \frac{\overline{n}}{T^2}\left[\left(\frac{\overline{u}}{\overline{u}+\overline{w}}\right)^2\sigma_w^2 + \left(\frac{\overline{w}}{\overline{u}+\overline{w}}\right)^2\sigma_u^2\right] . \tag{28}$$

The "master equation" (28) was presented the first time in [2]. Due to the relations

$$\frac{\overline{w}}{\overline{u}+\overline{w}} = q_w \quad (29) \qquad \frac{\overline{u}}{\overline{u}+\overline{w}} = q_u \quad (30) \qquad \overline{v} := \frac{\overline{n}}{T} \quad (31)$$

equivalent presentations are

$$Var\ q = \frac{\overline{v}}{T}\left(q_u^2\sigma_w^2 + q_w^2\sigma_u^2\right) = \frac{1}{T}\frac{q_u^2 q_w^2}{\overline{v}}\left(\mathrm{v}_u^2 + \mathrm{v}_w^2\right) = \frac{q_u^2 q_w^2}{\overline{n}}\left(\mathrm{v}_u^2 + \mathrm{v}_w^2\right)\ . \quad (32\ a,b,c)$$

[2]. It can be stated that the variance formulas eqs. (27) and (28) are exactly valid in the limit of very high crossing numbers n . Since n does not necessarily appear as an intrinsic system parameter in eq. (28), what is evident by the first of the eqs. (32), it can be concluded that formula (28) is also valid for arbitrary values of n at least as far as they meet the condition (17). This is also supported by the fact that the concepts of mean and variance are not bound on a *particular* distribution as for example $N(\mu,\sigma^2)$. For application of eqs. (32 a,b,c) in measurement practice, instead of the expectations $\overline{n}$, $\overline{v}$, $\overline{u}$, $\overline{w}$, σ_u and σ_W , the corresponding estimators evaluated from the vectors u_i and w_i are to be inserted. These are for $\overline{n}$ the observed crossing number n in T , for $\overline{v}$, the ratio n/T and for $\overline{u}$ and $\overline{w}$ the average of the u_i and w_i respectively. In addition the σ-quantities are to be replaced by the corresponding standard deviations [1]. The parameters q_u and q_w ($q_u+q_w:=1$) and T are constants to be chosen according to the type of measurement task.

5. CONCLUSIONS

It can be demonstrated that the variance of the partition which is imposed by a fixed level on the intsantaneous values of a stochastically varying signal like an environmental sound pressure level can be derived with transparency by aid of the Central Limit Theorem of Statistics. Although this principally implies an approximation it has the advantage to allow an explicite treatment of the appropriate directly observable parameters of the signal's microstatistics. It can be concluded that the highly simple and, as to be expected, completely symmetric final result is valid in general.

7. REFERENCES

[1] R. V. Hogg, E. A. Tanis, "Probability and Statistical Inference". Macmillan, New York, London, 1989.

[2] A. Heiss, "Variance of the total level crossing time of a continuous random signal and measuring accuracy of community noise quantiles". Proc. 6th Int. FASE-Congress, Zürich (Switzerland), 1992, pp. 343-346.

[3] A. Heiss, "Measurement Accuracy of Quantiles Determined for Continuous Random Signals and Application to Selective Assessment of Environmental Noise". Proc. of the Institute of Acoustics, UK, Euronoise '92, London 1992, Vol. 14, Part 4, pp 445-452.

[4] R. Goodman, "Intoduction to Stochastic Models". Benjamin Cummings, Menlo Park Calif., USA, 1988.

[5] A. Heiss, "Varianz des zeitlichen Geräuschmittelwertes stochastischer Folgen kurzer Schallereignisse" (Variance of the mean sound pressure level of random short event time series (in German language)). Proc. Fortschritte der Akustik, DAGA 90, Vienna (Austria) April 1990, pp. 385 - 388.