Audio Engineering Society

Convention Paper

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

DIFFERENCE LEVEL
An objective audio parameter

Serge Smirnoff

SoundExpert, Moscow, Russia
ssmirnoff@soundexpert.info

ABSTRACT

This paper describes the objective parameter, called "Difference Level" that could be considered either as an extension of THD for non-periodic signals, or as one of the estimations of widely used difference signal. It could be used for instrumental measurements of signal degradation in various audio circuits and for psycho-acoustic research. Infinite grade impairment scale and corresponding method for measurements of perceived audio quality, based on this parameter is also proposed.

1. INTRODUCTION

Digital processing and transmission of audio signals are widely practiced now. One of the consequences is decreased significance of objective audio parameters and, on the contrary, increased importance of subjective evaluation of audio quality. The classical set of objective parameters such as Total Harmonic Distortion (THD), Intermodulation Distortion (IMD), Signal-to-Noise Ratio, Frequency response and so on [1], was never ideal even for analog circuits. It corresponds with perceived quality to some extent only. Digital processing and psycho-acoustic algorithms in particular make that set and the measurement methods almost senseless. While new objective parameters are still under construction, the increased importance of subjective evaluation seems quite reasonable and effective.

On the other hand, reliable and repetitive subjective expertise is a very expensive, technically and organizationally complicated action. That's, probably, the main reason of ever increasing technical and marketing speculations around audio quality of consumer and professional devises so typical earlier perhaps for Hi-End area only. This dark side of subjective methods domination stimulates the search for new objective audio parameters even not perfect but suitable for use in Digital Millennium. Recently appeared PEAQ method [2], being a sophisticated combination of objective and subjective approaches, is a promising one. However high complexity and absence of independent realizations along with narrow implementation area make the wide spread of PEAQ method difficult. The more so as the potential of pure instrumental objective measurements is not exhausted yet.

2. DEFINITION

Let $X=[x_1,x_2...x_n]$ be an input discrete signal for some tested object and $Y=[y_1,y_2...y_n]$ be an output one. Without loss of generality it can be assumed that signals X and Y have the same sample rate and time aligned, i.e. every sample x_i maps to y_i. Then for any $N\in\mathbb{N}$ consistent samples $x_1,x_2...x_n$ and $y_1,y_2...y_n$ the correlation coefficient $\rho(X,Y)$ can be computed. In our case, it will show how much the shape of signal Y differs from the shape of signal X. As it will be shown further, the use of expression $\sqrt{1-|\rho|}$ instead of pure ρ is more practical (habitual), so the new parameter Df could be defined as follows:

$$Df(X,Y)=\sqrt{1-|\rho(X,Y)|}, \qquad (1)$$

where ρ is correlation coefficient.

Thus Df can take values from 0 to 1 depending on shape similarity of signals X and Y. Df can be expressed both in percent and decibels:

$$Df\,[\%] = Df\cdot 100\,\%; \quad Df\,[dB] = 20\,\lg Df \qquad (2)$$

By definition,

- $Df \equiv 0 \equiv 0\,[\%] \equiv -\infty\,[dB]$, if sound signals X and Y have exactly the same shape,
- $Df \equiv 1 \equiv 100\,[\%] \equiv 0\,[dB]$, if signal shapes are completely different.

As Df varies greatly in practice the use of logarithmic units is convenient. For example, DC offset removal in audio editor (16 bit mode) results in $Df \approx 0.003\%$ *(-90.5 dB)* and for psycho-acoustic coding-decoding $Df \approx 50\%$ *(- 6.0 dB)*.

3. FEATURES

3.1. Difference level does not depend on amplitudes and DC offsets of both signals.

The feature follows from correlation coefficient characteristics. This makes Df sensitive to the change of signal shape only while amplitudes and DC offsets are not taken into account. This is convenient for practical calculations of Df and very reasonable, because exactly the shape of a signal contains the whole information about sound image.

3.2. The physical essence of Difference level is normalized level of difference signal.

Indeed we may assume additionally that signals X and Y have no DC offsets (a) and RMS level of X equals to that of Y (b). Then the expression for Df can be transformed as follows (see annex 1):

$$Df(X,Y)=\sqrt{1-|\rho(X,Y)|} \Rightarrow \sqrt{\frac{\sum(y-x)^2}{2\sum x^2}} = \frac{P_{(Y-X)}^{RMS}}{P_X^{RMS}\sqrt{2}} \qquad (3)$$

We see that Df being a dimensionless value is the ratio of RMS levels of two signals – difference *(Y-X)* and input *(X)* ones. Thanks to this feature it is possible to compute Df for any signals represented as computer files in any audio editor. Beforehand the signals have to be equaled by RMS level, time aligned and their DC offsets have to be removed:

$$Df(X,Y) \approx P_{(Y-X)}^{RMS} - P_X^{RMS} - 3.01\,[dB] \qquad (4)$$

This feature gave the name to the parameter Df – "Level of difference signal" or just "Difference Level".

3.3. Difference level depends on the type of testing (input) signal.

Most interesting input signals are: sinusoidal, white noise, various real sound signals or their probabilistic models. As an example, Table 1 contains Df values for three input signals and two tested objects (software operations): 5dB increase of signal amplitude in audio editor (16 bit mode) and mp3 coding-decoding (128 kbit/s).

	Sin Wave (1KHz, -6 dB)	White Noise (-6 dB)	Harpsichord (-6 dB)
+5 dB gain	-93.87 dB	-89.15 dB	-75.97 dB
Mp3 128 kbit/s	-25.95 dB	-4.21 dB	-16.32 dB

Table 1: Df values for three input signals and two tested objects

Obviously, indicating some *Df* value it is necessary to point out the type of testing signal used.

3.3.1. Sinusoidal input signal and Total harmonic distortion (THD).

In case of sinusoidal input signal the expression for *Df* can be transformed as follows (see annex 3):

$$Df = \sqrt{1 - \frac{1}{\sqrt{THD^2 + 1}}} \quad \text{and} \quad THD = \sqrt{\frac{1}{(1 - Df^2)^2} - 1} \qquad (5)$$

The diagram is shown in Figure 1.

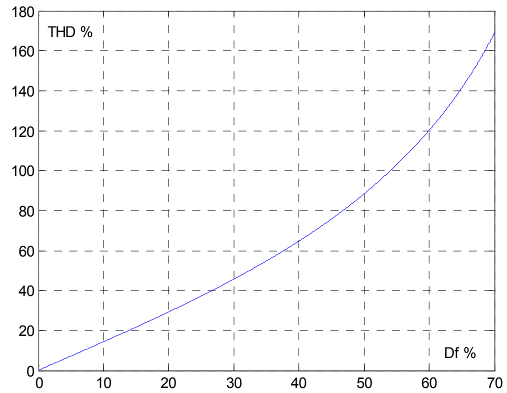

Figure 1: *THD – Df* diagram, %

It follows from the function properties that:

- *THD=100%* is equivalent to *Df≈54.1%* because the values *Df=0* and *Df=1* correspond by definition with extreme, theoretically possible cases and in this sense are absolute, while the value *THD=1* was chosen conventionally assuming the equality of the RMS voltages of the divisible harmonics to that of the fundamental. That was grounded and convenient as in practice THD never exceeds the level of 100%.

- For small *Df* and *THD* (*Df* $\leq$ *7%, THD* $\leq$*10%)* the first order Maclaurin polynomial approximation can be used:

$$Df = \frac{THD}{\sqrt{2}} \quad \text{and} \quad THD = Df\sqrt{2} \qquad (6)$$

with the following computational error ε, which tends to zero as THD→0:

Df, %	*THD*, %	ε, %
11.5	16.3	1
1	1.41	0.0075

Table 2: *THD – Df* simple conversion error.

The use of logarithmic values makes those formulas even simpler:

$$Df = THD - 3.01 \text{ [dB]} \quad \text{and} \quad THD = Df + 3.01 \text{ [dB]}. \qquad (7)$$

In fact, for small values (*Df* $\leq$ *-23.1 dB, THD* $\leq$ *-20 dB)* THD is 3 dB greater than *Df.* This is clearly seen from the above diagram plotted in logarithmic units (Fig. 2).

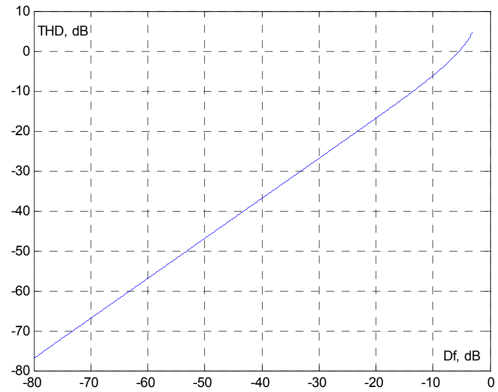

Figure 2: *THD – Df* diagram, dB

Thus, THD could be considered as a special case of Difference level when testing signal is sinusoidal. Let us remark that relation between *Df* and THD is true only for the tested objects that do not introduce additional noise. In general case Difference level corresponds to (THD+Noise). Similarly, in case of two sinusoidal signals Difference level corresponds to (THD+Noise+IMD).

3.3.2. Other input signals.

Generally talking, Difference level makes it possible to analyze an object under test in details simply modifying and combining input signals. The measurement procedure always stays the same and results can be compared with each other. White noise has to be mentioned specially. Its spectral complexity makes it to be the most severe test for any audio equipment. White noise is usually transferred through tested object with significantly less accuracy than harmonic signal. As an example, analog loop-back of certain semi-professional PC sound card under sinusoidal input signal shows the value *Df = -68 dB* which is equivalent to *THD = -65 dB (0.02%)* and close to the specification. After replacement of harmonic signal by white noise Difference level drops to *Df = -24 dB* which is conventionally equivalent to *THD = 8.8%.* Probably it can be assumed that a tested device under white noise introduces all possible types of distortion. Therefore parameter Df_{WN} could be helpful for end users of audio equipment as it indicates the equipment potential under worst case conditions while THD shows it under most favorable ones and useful mostly for developers.

Non-typical use of Difference level.

Due to above dependency of Difference level upon the type of input signal the parameter can be used in an unusual way. The example below, being more educational than practical, helps to disclose its additional possibilities. We will use it for investigation

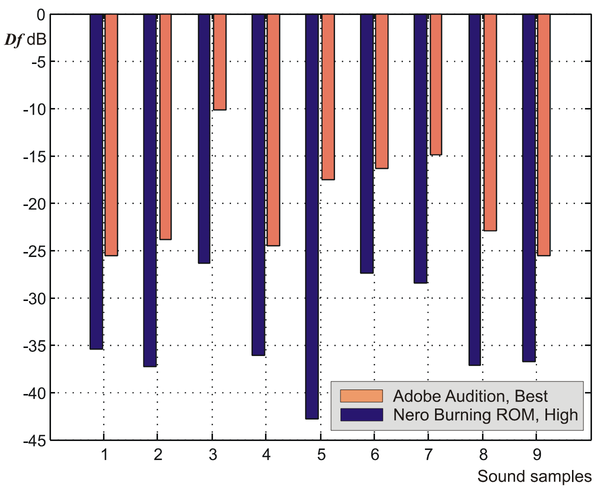

Figure 3: *Df* values of nine different sound samples, mp3-coded by two codecs.

of psycho-acoustic algorithms behavior. If we take several different input signals, then a tested codec will show different *Df* values for each of them. For instance, Figure 3 displays *Df* values for two codecs (mp3 128 kbit/s) and nine different sound samples from annex 3. It's obvious that our sound samples were modified by the codecs to different extent according to their psycho-acoustic models and other features. Consequently such set of *Df* values could be a kind of codec's "individual signature". Table 3 of annex 3 contains *Df* values for 10 codecs and 9 sound samples mentioned above. Now we can find similarities of the codec's "individual signatures" with the help of cluster analysis. Resulting dendrogram (Fig. 4) shows how the tested codecs have grouped with each other.

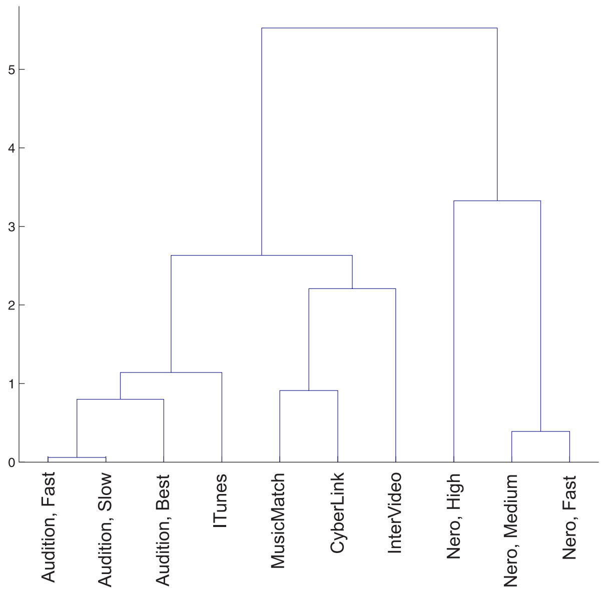

Figure 4: Codecs divided into groups according to their similarity.

The less the height of Π-shaped lines the more similar the connected objects. Looking at the dendrogram we can say with high probability that:

- Nero Burning ROM and Adobe Audition use for mp3 coding completely different coders (or completely different versions of the same coder);
- iTunes and Adobe Audition use for mp3 coding the same mp3 coder (perhaps they differ with some

fine settings, which are locked for the users of iTunes);
- MusicMatch and CyberLink Power Pack use the same mp3 coder;
- mp3 coder from InterVideo XPack is not like any other from the used set.

Probability of conclusions could be increased by selecting more appropriate sound samples, increasing the number of samples (codecs) and optimizing the procedure of cluster analyses. But it's already clear that this method could be helpful when you need to determine the type of some coder (or another tested object) by comparing it with the known ones or for example, to trace changes of a coder engine from version to version.

3.4. Difference level can be computed for signals both in time and frequency domains.

Using Parseval's theorem it is not hard to prove that:

$$Df(X,Y) \equiv Df(\mathscr{F}[X], \mathscr{F}[Y]), \tag{8}$$

where $\mathscr{F}[X]$ and $\mathscr{F}[Y]$ are complex spectrums of corresponding signals.

3.5. Resampling and time alignment.

Difference level is not affected by identical upsampling of X and Y signals if upsampling itself is precise enough, i.e. if Df value of the algorithm is substantially less than $Df(X,Y)$. But it is very sensitive to time alignment errors, which are caused by:

- constant time or phase shift;
- time stretch/shrink.

This is a point where Difference level could be seriously criticized. Fortunately digital processing is time accurate in most cases (psycho-acoustic coding in particular). Unfortunately in most cases digital-to-analog and analog-to-digital conversions are not. Computationally efficient and precision controlled algorithm of resampling has to be developed. In case of exact time alignment Df value tends to its minimum and auditory perceived correlation between difference and input signals is the lowest.

4. RELATION TO SUBJECTIVE MEASUREMENTS.

Possibility of using real sound samples for computation of Difference level makes it possible to study relationship between the new parameter and subjective estimates directly – by comparing Df values with corresponding subjective grades. Please note, that the results below are the first of the kind and need more research but promising and interesting enough to be presented in this paper though.

Figures 5 and 6 show subjective ratings and corresponding difference levels for two sound excerpts and six tested objects – psycho-acoustic algorithms (see annex 4 for the listening test details). Df values for each sample were centered and scaled to standard deviations of the subjective grades beforehand.

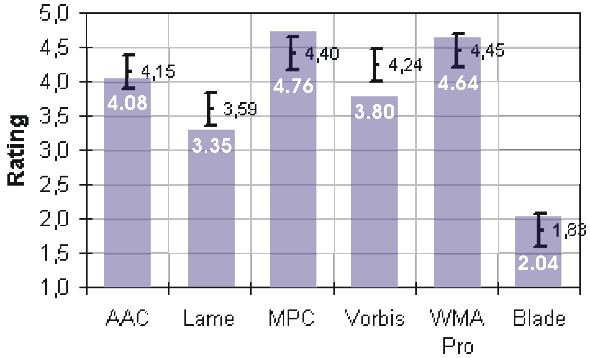

Figure 5: Subjective ratings and corresponding Df values (centered and scaled) for "Layla" sound sample. $\rho = 0.95$

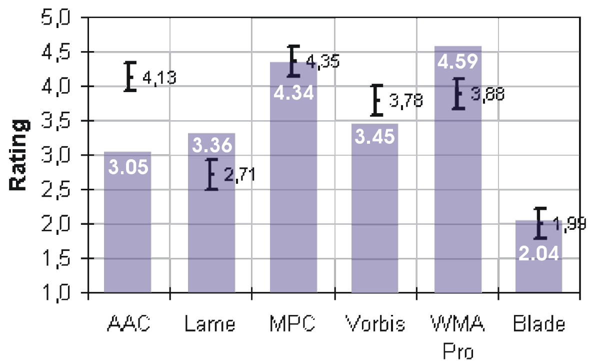

Figure 6: Subjective ratings and corresponding Df values (centered and scaled) for "Waiting" sound sample. $\rho = 0.74$

Taking into account that psycho-acoustic coding is the hardest test possible for any objective measurements, we could say that the above correlation is high enough. Nevertheless, today we all are clever enough not to measure perceptual audio quality with any simple instrumental parameter or even a set of them. Difference level is not an exception. In some cases it can be very deceptive because it shows only the amount (quantity) of signal shape change and does not take into account the type (quality) of that change. However, the above correlation increases substantially when the type of differences is fixed. Good example of the case is a sound excerpt mixed with some outside signal in various proportions. *Df* values for that set changes from sample to sample while the type of the change stays the same. Relation between Difference level and auditory perception in such a case can be studied both by comparing *Df* values with corresponding subjective grades according to some impairment scale and in terms of "Just Noticeable Difference". Also various combinations of test-outside signals could be observed. Outside signal could be natural or artificial with time/spectral structure adapted or not to that of the test one. Such different situations as mixing of some noise to a signal and mixing of additional instrument to a composition can be analyzed with a single approach and common objective parameter. In both cases Difference level will show the RMS level of outside signal relative to that of the main one. And $Df = 1$ *(0 dB)* will indicate the equality of those levels.

So called "difference signal" widely used in various measurement schemes can be considered as an example of outside signal. It is usually mixed with "reference signal" in some proportion. The resulting signal, usually being an output of some "device under test" is often called "signal under test". The psychometric experiment below deals with this kind of input and outside signals.

There were three devices under test used:

- ADPCM 434 kbit/s codec
- MP3 320 kbit/s codec
- WMA 320 kbit/s codec

and two reference sound samples from SQAM disk:

- French male speech
- Harpsichord

For each device and sound sample a set of five artificial test items were produced by subtracting gradually increasing amount of reference signal from output one with subsequent normalization of resulting signals:

$$Z_n = Y - K_n * X \qquad (9)$$

where K_n increasing linearly from $K_1=0$ and chosen so that sound artifacts in resulting signals Z_n would be audible for ordinary listeners.

As it was showed in [3] additional filtering of the resulting signal is necessary for elimination of the frequency components, absent in output signal and thus appeared in resulting signal after subtraction. Such filtering was performed by FIR filter with the magnitude response vector, calculated according to slightly modified algorithm from [3].

As the main idea of the experiment was to analyze the interrelationship of subjective and objective parameters in principle and sound artifacts in all test items were easily audible a simplified two stimulus hidden reference double-blind listening test was performed. Test items from different codecs were mixed and randomized. Three listeners graded basic audio quality using their home stereos and headphones with all enhancements/equalizations turned off. Each of them repeated the whole test, consisting of 60 double stimulus test items, twice. So each subjective grade was calculated on the basis of 12 observations. Listeners were advised to pass an advance training in order to remember the reference samples and to become familiar with 5-grade impairment scale. Raw results of the listening test could be found in annex 5.

Figure 7 shows mean values and Student's confidence intervals at corresponding Difference levels. We see that psychometric functions are:

- monotonic for all samples and codecs used,
- well consistent with second order curve.

NOTE 1. Monotonic character of psychometric functions was also confirmed in [3].

NOTE 2. The consistency with second order curve is not ideal even in case of perfectly designed listening tests. Non-linear nature of human hearing and masking thresholds makes perception of gradually unmasked artifacts uneven. The level of the consistency depends on the type of reference/difference signals.

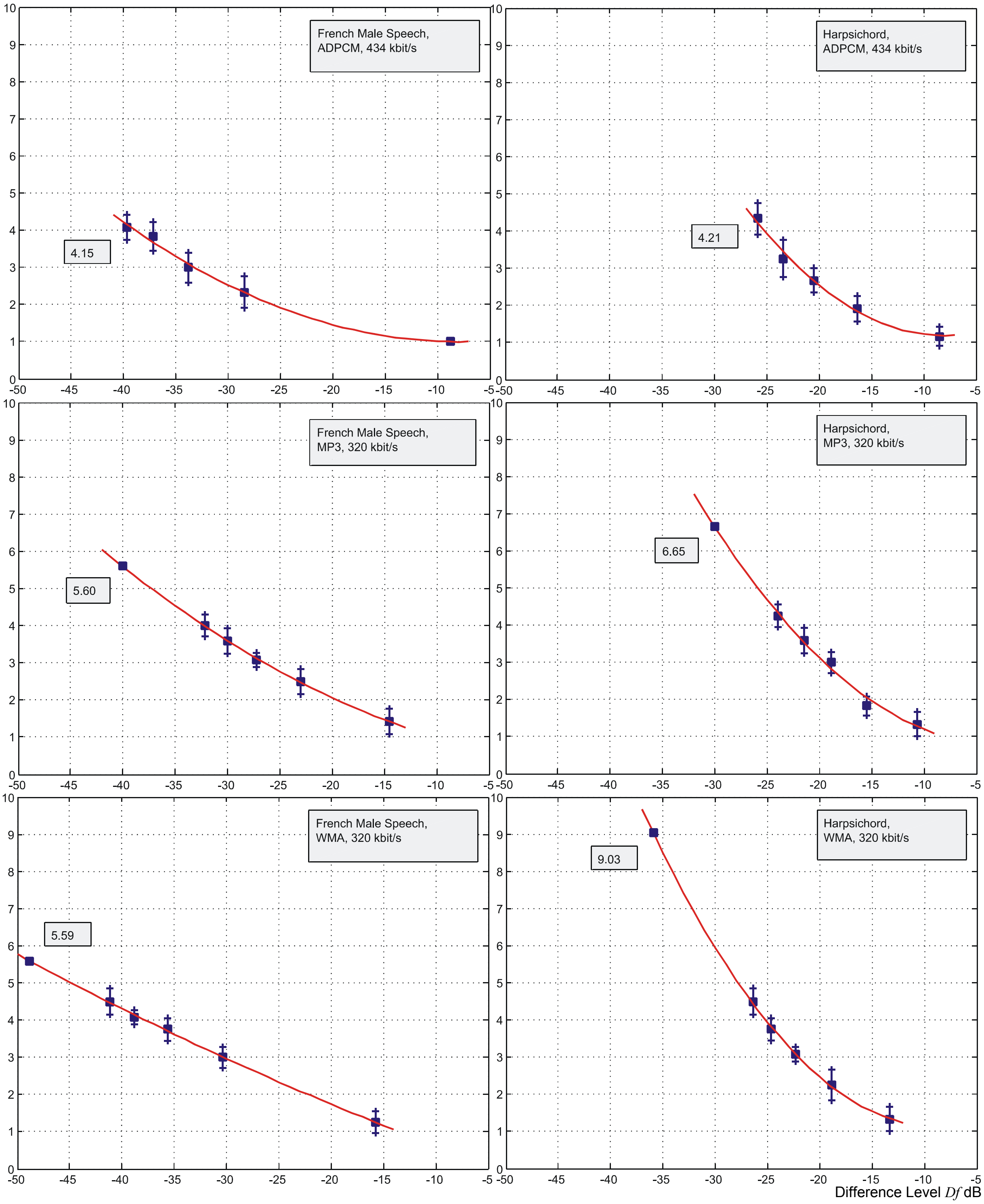

Figure 7: Subjective scores and corresponding psychometric functions.

NOTE 3. Codecs and sound samples from this listening test differ from each other essentially but still show similar psychometric functions. In some informal tests those functions had negative curvature but in most cases they look like the ones above – 2nd order with positive curvature or linear.

Infinite Grade Impairment Scale.

It seems reasonable to approximate such functions with 2nd order polynomial. The analytical curves are shown on the diagram as well. Now it's easy to compute imaginary or virtual subjective grades of natural output signals with main Difference levels. These predicted points are shown without confidence intervals but the latter can be found by means of Monte Carlo simulation. Factors affecting the final accuracy of interpolation have to be researched. However some preliminary recommendations for improving accuracy are following:

- Increasing the number of testing points. First experiments show that 3 points are absolute minimum and 5 to 10 points are optimal for simple design of listening tests. The design itself affects the accuracy to less extent.
- Avoiding test items close to the edges of 5-grade impairment scale. Mean values between *2* and *4.5* are OK.

Analytically computed random values of virtual ratings can be further analyzed by ANOVA as usual.

In order to "grade" the predicted virtual points, standard 5-grade impairment scale was continued to infinity and thus transformed into infinite grade impairment scale (IGIS).

NOTE 1. Subjective grades of artificial test items are less dependent on listening test design, because sound artifacts are audible by majority of ordinary listeners.

NOTE 2. Virtual part of IGIS is non-linear as like the real one and completely depends on the latter.

NOTE 3. It is clearly seen from the diagram that the highest ADPCM grades, belonging to the natural test items, are also well fitted to the curves and could be predicted by remaining four "artificial" grades. Therefore real and virtual parts of IGIS join to each other smoothly.

Direct testing and grading of the natural output signals is far beyond the possibilities of our simple design listening test. Moreover, testing items with such small impairments would cause a problem even for perfectly designed listening test. Therefore IGIS could be useful for grading very small impairments or the moderate ones as it reduces listening test expenses.

Infinite Grade Measurement Scheme.

Measurement scheme based on particular relationship between Difference level and subjective grades is shown schematically in Figure 8.

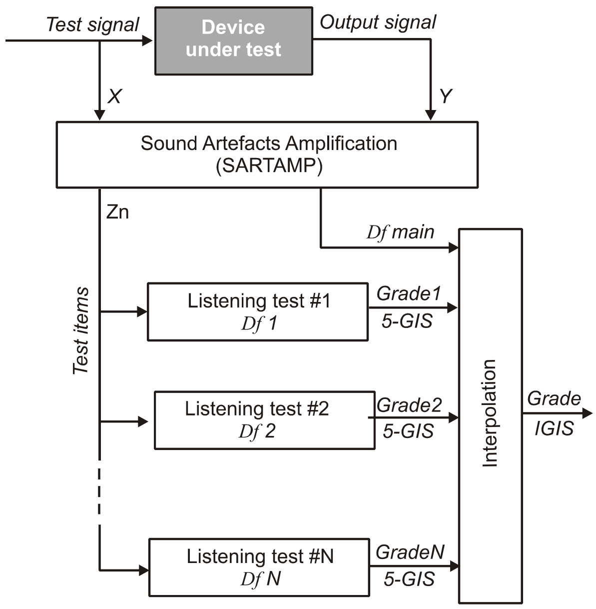

Figure 8: Infinite Grade Measurement Scheme.

Device under test could be of any kind – software or hardware, digital or analog. The latter needs high precision DA/AD converters and effective time alignment mechanism. SARTAMP aimed to perform that time alignment of X and Y, their subtraction in various proportions and filtering of resulting signals Zn. Listening tests could be of simple design with participation of ordinary listeners. Interpolation block calculates unknown virtual subjective grade with confidence intervals.

This measurement scheme can be considered as further development of "coding margin" method described in [3]. In this paper perceptual quality of output signal with small impairments is estimated by the difference signal gain required for the artifacts become audible in resulting mix with input signal. In our case the coding margin can be approximately expressed as:

$$g(dB) = Df(5\text{-}grade) - Df(main) \tag{10}$$

where *Df(5-grade)* is the *Df*-coordinate of the 5-grade curve point.

It is clearly seen from Figure 7 that virtual subjective grade in terms of [3] is a map of coding margin onto infinite grade impairment scale by means of psychometric function (+5, to be correct).

On the other hand, this subjective-objective measurement scheme is similar to already mentioned PEAQ method. The main difference is subjective part, which is formalized and incorporated inside the PEAQ scheme and, on the contrary, exists in the form of specially designed listening tests in IGIS scheme. In fact, virtual subjective grade is a combination of objective measurement of signal change by Difference level and subjective estimation of that change by means of psychometric function in listening tests.

5. SUMMARY

As many other audio parameters Difference level derives from comparison of input *(X)* and output *(Y)* signals of some device under test. It indicates the level of the difference *(Y-X)* signal relative to the input one. The application area of the parameter is quantitative estimation of signal shape modification caused by some signal processing or transmission. Both input signal and device under test can be of any kind. In case of sinusoidal input signal Difference level is almost equal to (THD+Noise). The use of other input signals – real or artificial helps to get more detailed and "real-world" objective characteristics of audio circuits. The name of the proposed parameter does not contain the word "distortion" intentionally. This word has some negative meaning and could mislead while the relationship between any objective parameters and perceived audio quality is still complicated and mostly undefined. The research of such relationship can be assisted by the Difference level as well, because it can be computed for various sound samples suitable for subjective testing. The experiment with specially prepared sound samples reveals that the relationship between Difference level and auditory perception in case of the same difference signal can be described by the 2nd order psychometric functions. Based on these functions infinite grade impairment scale is proposed. It could be useful for sound quality assessment of small impairments. Accuracy of this method has to be researched. Another problem that needs additional research is time alignment, which affects Difference level significantly.

6. ACKNOWLEDGEMENTS

The author of this paper would like to thank the following people:

- Jacco Dekkers for his valuable comments and advices;
- Andrey Merzlyakov, Vasily Vasilenko and Dmitry Timofeev for participation in listening tests.

7. REFERENCES

[1] CABOT R.C, "Fundamentals of Modern Audio Measurement." J. Audio Eng. Soc., vol. 47, (1999 September)

[2] ITU-R, 2001. Method for objective measurements of perceived audio quality. ITU-R Recommendation BS.1387-1.

[3] FEITEN B, "Measuring the Coding Margin of Perceptual Codecs with The Difference Signal" presented at the AES102th Convention, Munich, Federal Republic of Germany, 1997 March, preprint 4417

ANNEX 1 $M(X)=M(Y)=0$ (a) (11)

$P_x = P_y \Rightarrow \sum x^2 = \sum y^2$ (b) (12)

$$Df(X,Y)=\sqrt{1-|\rho(X,Y)|};$$

$$1-|\rho(X,Y)|=1-\left|\frac{M(XY)}{\sqrt{M(X^2)}\sqrt{M(Y^2)}}\right|=1-\sqrt{\frac{M(XY)M(XY)}{M(X^2)M(Y^2)}}=1-\sqrt{\frac{\left[\sum(xy)\right]^2}{\left[\sum(x^2)\right]^2}}=1-\frac{\sum(xy)}{\sum x^2}=\frac{\sum x^2-\sum(xy)}{\sum x^2}=\frac{\sum(x^2-xy)}{\sum x^2}= \quad (13)$$

$$\frac{2\sum(x^2-xy)}{2\sum x^2}=\frac{\sum(2x^2-2xy)}{2\sum x^2}\Rightarrow\frac{\sum(x^2+y^2-2xy)}{2\sum x^2}=\frac{\sum(y-x)^2}{2\sum x^2};$$

$$Df(X,Y)=\sqrt{1-|\rho(X,Y)|}=\sqrt{\frac{\sum(y-x)^2}{2\sum x^2}}.$$

ANNEX 2 For sinusoidal signals we have:

$X=U_1\sin(t+\varphi_1)$; $\quad Y=U_1\sin(t+\varphi_1)+U_2\sin(2t+\varphi_2)+...+U_n\sin(nt+\varphi_n)=\sum_{n=1}^{\infty}U_n\sin(nt+\varphi_n)$ (14)

Taking into account that:

$M(\sin^2 t)=\frac{1}{2}$, $\quad M(\sin t)=M(\cos t)=0$, $\quad \sin\alpha\sin\beta=\frac{\cos(\alpha-\beta)}{2}-\frac{\cos(\alpha+\beta)}{2}$, (15)

we can find each of correlation coefficient factors

$$\rho=\frac{M(XY)}{\sqrt{M(X^2)}\sqrt{M(Y^2)}}. \quad (16)$$

$$\sqrt{M(X^2)}=\sqrt{M(U_1^2\sin^2(t+\varphi_1))}=\frac{U_1}{\sqrt{2}}. \quad (17)$$

$M(XY)=M\left[U_1\sin(t+\varphi_1)\cdot\sum_{n=1}^{\infty}U_n\sin(nt+\varphi_n)\right]=M\left[U_1^2\sin^2(t+\varphi_1)+S_1\right]$, where (18)

$S_1=U_1U_2\sin(t+\varphi_1)\sin(2t+\varphi_2)+U_1U_3\sin(t+\varphi_1)\sin(3t+\varphi_3)+...+U_1U_n\sin(t+\varphi_1)\sin(nt+\varphi_n)$, (19)

for which *M[S₁] = 0* according to (15). Thus we obtain

$$M(XY)=M\left[U_1^2\sin^2(t+\varphi_1)\right]=\frac{U_1^2}{2}. \quad (20)$$

$\sqrt{M(Y^2)}=\sqrt{M\left[\sum_{n=1}^{\infty}U_n\sin(nt+\varphi_n)\right]^2}=\sqrt{M\left[\sum_{n=1}^{\infty}U_n^2\sin^2(nt+\varphi_n)+2S_2\right]}$, where (21)

$S_2=\sin(t+\varphi_1)\sin(2t+\varphi_2)+\sin(t+\varphi_1)\sin(3t+\varphi_3)+...+\sin((n-1)t+\varphi_{n-1})\sin(nt+\varphi_n)$ (22)

for which $M[S_2] = 0$ according to (15). Thus we obtain

$$\sqrt{M(Y^2)}=\sqrt{M\left[\sum_{n=1}^{\infty}U_n^2\sin^2(nt+\varphi_n)\right]}=\sqrt{\sum_{n=1}^{\infty}M\left[U_n^2\sin^2(nt+\varphi_n)\right]}=\sqrt{\sum_{n=1}^{\infty}U_n^2\cdot\frac{1}{2}}=\frac{\sqrt{\sum_{n=1}^{\infty}U_n^2}}{\sqrt{2}}. \quad (23)$$

Now considering $THD=\frac{\sqrt{U_2^2+U_3^2+...+U_n^2}}{U_1}=\frac{\sqrt{\sum_{n=2}^{\infty}U_n^2}}{U_1}$, we have (24)

$$\rho=\frac{M(XY)}{\sqrt{M(X^2)}\sqrt{M(Y^2)}}=\frac{\frac{U_1^2}{2}}{\frac{U_1}{\sqrt{2}}\frac{\sqrt{\sum_{n=1}^{\infty}U_n^2}}{\sqrt{2}}}=\frac{U_1}{\sqrt{\sum_{n=1}^{\infty}U_n^2}}=\sqrt[-1]{\frac{\sum_{n=1}^{\infty}U_n^2}{U_1^2}}=\sqrt[-1]{\frac{\sum_{n=2}^{\infty}U_n^2}{U_1^2}+\frac{U_1^2}{U_1^2}}=\sqrt[-1]{THD^2+1}=\frac{1}{\sqrt{THD^2+1}} \quad (25)$$

and finally $\quad Df=\sqrt{1-|\rho|}=\sqrt{1-\frac{1}{\sqrt{THD^2+1}}}.$ (26)

ANNEX 3

Here and further 9 sound samples from the project SoundExpert (http://www.soundexpert.info) were used:

1. Bass (SQAM)
2. Castanets (SQAM)
3. French Male Speech (SQAM)
4. Glockenspiel (SQAM)
5. Harpsichord (SQAM)
6. J.S.Bach, "Oster-Oratorium, BWV 24"
7. Lo-Fi (home) Compact Cassette Recording
8. Mike Oldfield, "Music From The Balcony"
9. Quartet (SQAM)

They were converted to “mono” beforehand in order to eliminate the influence of coder settings concerning the coding of stereo image. In some programs this settings are locked for end user.

Coders used:

1. Ahead Nero Burning Rom 6.0.0.23, mp3 128 kbit/s, High
2. Ahead Nero Burning Rom 6.0.0.23, mp3 128 kbit/s, Medium
3. Ahead Nero Burning Rom 6.0.0.23, mp3 128 kbit/s, Fast
4. Adobe Audition 1.0, mp3 128 kbit/s, Best
5. Adobe Audition 1.0, mp3 128 kbit/s, Fast
6. Adobe Audition 1.0, mp3 128 kbit/s, Slow
7. iTunes 4.1.1.54, mp3 Good Quality, 128 kbit/s
8. MUSICMATCH Jukebox 8.10.2017CJJ, mp3 128 kbit/s,
9. CyberLink MPEG Layer-3 Encoder 1.00 from CyberLink Power Pack, mp3 128 kbit/s
10. InterVideo mp3 Encoder from InterVideo XPack (DVD and MP3), mp3 128 kbit/s

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9
Nero, High	-25.59	-23.89	-10.26	-24.58	-17.61	-16.34	-14.91	-22.94	-25.54
Nero, Medium	-29.06	-30.85	-13.54	-27.13	-34.211	-19.44	-18.23	-28.60	-29.54
Nero, Fast	-35.46	-37.30	-26.41	-36.18	-42.83	-27.40	-28.54	-37.17	-36.72
Audition, Best	-36.40	-38.03	-24.86	-37.22	-42.12	-25.11	-29.53	-36.22	-37.54
Audition, Fast	-36.47	-38.08	-24.87	-37.27	-42.21	-25.25	-29.57	-36.41	-37.56
Audition, Slow	-38.47	-37.73	-23.23	-34.25	-40.98	-26.77	-25.83	-35.06	-37.85
iTunes	-44.12	-45.76	-24.02	-36.36	-36.71	-22.30	-28.81	-34.90	-44.26
MusicMatch	-45.84	-46.85	-24.29	-38.32	-34.76	-21.91	-28.08	-32.96	-41.19
CyberLink	-39.53	-40.28	-19.76	-33.51	-34.51	-20.72	-22.40	-32.05	-38.96
Intervideo	-30.27	-31.80	-13.87	-27.92	-34.52	-19.89	-18.55	-29.23	-30.52

Table 3: Difference levels for the above 9 sound samples and 10 coders.

ANNEX 4

«128kbps Extension public listening test» was conducted by Roberto Amorim with the participants of Hydrogen Audio Forum. (http://www.hydrogenaudio.org).

Detailed information about methods and results of the test can be found at http://www.rjamorim.com/test/128extension/presentation.html.

The following coders were under test:

1. AAC. Apple QuickTime 6.3 MP4 encoder 128kbps, high quality
2. BladeEnc 0.94.2, -128
3. LAME MP3 Encoder 3.90.3, --alt-preset 128
4. MPC, Musepack 1.14, --quality 4 –xlevel
5. Ogg Vorbis post-1.0 CVS, -q 4.25
6. Windows Media Audio v9 PRO, bitrate-managed 2-pass VBR 128kbps

In this paper the results for two sound samples were used:

1. Eric Clapton, Layla (unplugged)
2. Green Day, Waiting (Album “Warning”)

ANNEX 5

		Testing Item 1	Testing Item 2	Testing Item 3	Testing Item 4	Testing Item 5
French Male Speech	**ADPCM 434 kbit/s -39.6938 dB**	-8.7575 dB	-28.4108 dB	-33.8470 dB	-37.1993 dB	-39.6938 dB
		111111111111	122222223333	223333333344	333444444445	344444444455
	MP3 320 kbit/s -40.0584 dB	-14.5704 dB	-22.9774 dB	-27.1980 dB	-30.0180 dB	-32.1395 dB
		111111122222	222222333333	333333333334	333334444444	344444444445
	WMA 320 kbit/s -48.8111 dB	-15.7888 dB	-30.3403 dB	-35.6183 dB	-38.8289 dB	-41.1527 dB
		111111111222	233333333334	333444444444	444444444445	444444555555
Harpsichord	**ADPCM 434 kbit/s -25.8267 dB**	-8.5047 dB	-16.3235 dB	-20.4824 dB	-23.4674 dB	-25.8267 dB
		111111111122	112222222223	222233333333	233333333445	444444444556
	MP3 320 kbit/s -30.0341	-10.6286 dB	-15.4869 dB	-18.8225 dB	-21.4581 dB	-23.9538 dB
		111111112222	112222222222	233333333334	333334444444	444444444555
	WMA 320 kbit/s -35.9174	-13.3450 dB	-18.8414 dB	-22.3567 dB	-24.6925 dB	-26.3798 dB
		111111112222	222222222234	333333333334	333444444444	444444555555

Table 4: Raw results of the listening test with corresponding Difference levels of testing items.