Analysis of Non Linear Distortion in Compression Drivers. **3998 (L4)**

Philippe Robineau and Rémi Vaucher, Nexo Distribution, Roissy CDG, France.

Presented at
the 98th Convention
1995 February 25 - 28
Paris

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

ANALYSIS OF NON-LINEARITIES IN COMPRESSION DRIVERS AND HORNS

Philippe ROBINEAU *Rémi VAUCHER*

NEXO 154 allée des Erables, BP 50107 95950 ROISSY CDG FRANCE

This paper will show that the main cause of distortion in compression drivers is the air non linearity that occurs in the horn and phase plug and not in the transducer itself. This non-linearity is mainly a result of the modulation of the celerity the sound wave by the acoustic velocity. Modeling has been made and successfully experimented by pre-correcting the audio signal in the digital domain.

I. Introduction

Non linearities in transducers are a well known fact. This phenomenon is particularly audible and bothersome if the transducer is a pressure chamber loudspeaker and in this case a high level of distortion (may be above 30%) arises in the high and mid-high frequencies. In most cases this distortion limits the use of the transducers which could be otherwise driven nearer to their mechanical and thermal limits.

Although magnetic and mechanical non linearities may also exits in horn compression drivers, as in mid range or bass loudspeakers [6] [7], the predominant distortion is induced by the non-linear behavior of the transmission medium (i.e. air) associated with high acoustical intensity.

Distortion having been analyzed, and then theoretically modeled as a non-linear transformation [P] applied to the acoustic signal x(t), it has been possible to find an inverse transformation $[P^{-1}]$ (i.e. such as $P[P^{-1}(x(t))]=x(t)$) to be applied to the electrical signal as a pre-correction.. The pre-correction algorithm has been tested in simulation and experiments, and then implemented in the digital domain to achieve a real time processing.

I.1. Brief technical background

Most transmission mediums are nonlinear as soon as the amplitudes of transmitted signals are large enough to induce deformations which can no longer be considered as small perturbations of the physical parameters of the said medium at its equilibrium state. This phenomenon can easily be seen in the case of the propagation of acoustic waves in a fluid medium.When air is the medium (supposed adiabatic), the fundamental formulas of the so-called « linear » acoustic waves propagation are deducted from :

- application of the two fundamental principles of the dynamic, mass conservation and kinetic quantity

$$\frac{\delta\rho}{\delta t} + \nabla \cdot (\rho \cdot \vec{u}) = 0 \qquad \text{[Eq. 1}$$

$$\frac{\delta(\rho\vec{u})}{\delta t} + \vec{grad}(p) = \vec{0} \qquad \text{[Eq. 2}$$

- linearisation of the adiabatic equation which links pressure and density in a (perfect) gas medium, $p = \rho^{\gamma}$ (where γ is the ratio of the specific heats.)

At the first order of the power series we have the linear acoustic equations. Validity of such a linearisation depends on the ratio between the 1st order coefficient (equal to c^2) and the further order coefficients. Classical non-linear acoustic equations are obtained taking in account the second order terms.

Considering high acoustic intensity transducers (such as compression drivers), it is known from both theoretical works ([2] p223 ; [1] p275) and experiments that :

- for a pure sine wave, nth Harmonic (H_n) is increased by (n-1)*3dB when acoustical power is doubled
- For an exponential horn of infinite length, H_n is increased by n*3 dB when the ratio between frequency of the emitted signal and the cut-off frequency of the horn is doubled, at the same intensity.

According to the previous statements, loudspeakers and cabinets manufacturers have often tried to reduce this distortion by physical means, but the work done in this area could not yielded satisfactory results. This inability to reduce distortion is mainly due to keep a certain trade off between contradictory requirements concerning the dimensions of the transducer and the associate horn :

- Obtaining an extended frequency response and maintaining directional properties at high frequencies implies small throat dimensions ;
- Achieving low distortion implies increasing the cross section of the throat of the transducer (near the pressure chamber) since for a given acoustical

power the sound pressure level conditioning the distortion is inversely proportional to the cross section.

- The cut off frequency, for a low distortion, must be as large as possible and that is incompatible with horns designed for having a wide frequency range.

II. Theoretical Model

II.1. Static and propagative distortion

Distortion in horn compression drivers is generally only viewed as a phenomenon of « air overload in the horn » [2]) : when applying the thermodynamic adiabatic law PV^{γ}= constant to large amplitudes - where 1st order approximation is no more valid - , the acoustic pressure created by compression or expansion of air does not follow a linear law. [1]. According to the law of energy conservation in an adiabatic flow :

$$\frac{T_0}{T} = 1 + \frac{\gamma - 1}{2} \cdot M^2$$

Laplace laws -> $\frac{P}{P_0} = \left(\frac{T_0}{T} \right)^{\frac{\gamma-1}{\gamma}}$

and so $\frac{P}{P_0} = \left(1 + \frac{\gamma - 1}{2} \cdot M^2 \right)^{\frac{\gamma-1}{\gamma}}$

This non-linearity (which we would call « static ») is nevertheless not the major cause of distortion. Introducing 2d order terms in equations 1 and 2 modifies the wave equations themselves, leading to another form of distortion which is « propagative », meaning that the celerity of the waves is depending on the acoustic velocity. According to previous works of the litterature on non-linear acoustics [3] an to our experiments, this distortion is in fact predominant in our field of interest.

II.2. Propagative distortion in case of a plane wave.

Considering the above fundamental equations applied to a plane wave on a Ox axis, vector u can now be expressed as the scalar variable V : the acoustic velocity at a point with abscissa x=0 is any function of time $V_0(t)$, and the velocity V(x,t) at any point with abscissa x is related to $V_0(t)$.

In case of linear approximation (1^{st} order, valid for low amplitudes) :

$V_{(x,t)}=V_0\ (t-x/C_0)$ which *[Eq. 3*

For a non-linear medium (2^{nd} order) :

$$V_{(x,t)} = V_0\left(t - \frac{x}{c_0 + \frac{\gamma+1}{2}V_0}\right) \qquad [Eq. 4$$

This means that the propagation celerity of points representing the signal amplitude is modulated by acoustic celerity, the celerity for a positive amplitude being greater than the celerity for a point with the same negative amplitude.

Comparing the equations (3) and (4) term by term shows up the basic principle of the correction since to a predetermined time-delay x/C_0 as a function of distance in the linear case there correspond a time delay :

$$\tau = \frac{x}{c_0 + \frac{\gamma+1}{2}V_0} \qquad [Eq. 5$$

if the non linearity of the propagation is taken into account.

Consequently it is exactly as if the acoustical wave at any point was distorted or modulated by variable delays dependent on the velocity V_0 at the origin. (Figures 4 & 5)

II.3. Application to propagation in horns.

We have seen a way (for a plane wave) to calculate τ function of the initial amplitude of the velocity V_0, and the length of the acoustic path. May this result be adapted to the case of acoustic horns ?

Considering a practical acoustic horn device such as shown in Figure 1, the acoustic path that has to be regarded extents from the pressure chamber output, which is then the true input (throat) of the horn. The first horn segment is formed by the channels of the phase plug, and on some length of the initial acoustical path of the horn the internal cross-section is relatively small (compared to the wave-length) so that the propagation can be considered as unidirectional, sphericity of the wave front being still neglictible.

As the non-linear effects are mostly produced where the cross-section is small (since depending on the acoustic intensity), the contribution to the overall distortion of the initial part of the horn is strongly predominant and one can reasonably consider that the unidirectional assumption will introduce neglictible errors in our concern.

The total distortion to be taken into consideration at the outlet from the horn (since non-linear distortion arising between the outlet from the horn and the listener can be neglected) may so be determined by integration allowing for the shape of the horn.([Eq. 6)

[Eq. 6

$$\tau(x) = \int_{Shape(x)} \frac{1}{C_0} \cdot \frac{dx}{1 + \frac{\gamma + 1}{2} \cdot \frac{V_0}{C_0}}$$

Specifically the shape of the horn may be characterized by expressing the variation in the amplitude $V_{(x)}$ of the velocity along the propagation path Ox. If the shape of the horn can be related to an analytical function S(x) expressing the variation in the cross-section of the horn along the Ox axis, it is possible to deduce a function V(x) characterizing the variation in the amplitude of the velocity along Ox.

In some cases the time delay function can itself be obtained in a relatively simple analytical form. $\tau(\xi_0, L, \chi_0)$, with $\xi_0 = \frac{\gamma+1}{2} \cdot \frac{v_0}{c_0}$

and x_0 which is dependent on the horn.

Subsequent equations are given for conical and exponential horns but the principle is valid for all horn whatever the shape they may have.

For a conical horn of length L and with a throat cross-section S_0 as defined by the equation :

[Eq. 7

$$S(x) = S_0 \cdot \left(\frac{x}{x_0} \right)^2$$

in which x_0 is a constant characteristic of the horn, the following equation is obtained :

[Eq. 8

$$\tau(\xi_0, L, x_0) = \frac{L}{c_0} - \xi_0 \cdot \frac{x_0}{c_0} Log\left(1 + \frac{L}{x(1+\xi_0)} \right)$$

where $\xi_0 = \frac{\gamma+1}{2} \cdot \frac{V_0}{c_0}$

For an exponential horn of length L and throat cross section S_0, as defined by the equation :

[Eq. 9

$$S(x) = S_0 \cdot \exp^{2ax}$$

where a=m/2 ; m, the flare of the horn, is a constant characteristic of the horn :

[Eq. 10

$$\tau(\xi_0, L, a) = \frac{L}{c_0} - \frac{1}{ac_0} Log\left(1 + \frac{1+\xi_0 \cdot \exp(-aL)}{1+\xi_0} \right)$$

Here the term « throat cross section » denotes the sum of the cross sections of the openings from the phase plugs in the diaphragm area.

II.4. Validation of the model.

The model considers the following statement (which have been verified by experiment).

Almost all distortion of the sound wave due to the non-linearity of the air occurs between the transducer and the outlet from the horn. Beyond this the wave propagates normally without significant further distortion.

The resulting distortion of the sound wave can be regarded as a variable phase shift (or delays) of certain parts of the sound wave, depending on the sound pressure level.

The goal of the model is therefore to predict how such distortion of the sound will evolve in order to apply compensating variable time delays and magnitude adjustment to the electrical signal driving the transducer.

II4.a.Simulation program (non-real time)

After having determined the laws ruling the propagative delay and static distortion, a program was developed to test the efficiency of those algorithms using real signals and a measurement environment based on :

- an arbitrary waveform generator (WAVETEK 75), programmable through a GPIB interface, including a 12 bits D/A converter with a programmable sampling rate up to 2 MHz.
- an FFT analyzer (ONO SOKKI CF 930).

The device under test was a compression driver TAD loaded with a NEXO HF horn powered by a gain calibrated amplifier. An equalizer was inserted between the generator and the amplifier to provide a flat linear transfer function between input electrical voltage and output acoustic velocity.

The program processed an input waveform file with a number N of data points and two processed data files were created :

- a file of the simulated non corrected acoustic waveform, for comparison with measurements.
- a file of the pre-corrected electrical input waveform.

The arrays of signal data points were derived by using a linear interpolation between data points computed at unequal time intervals, which were derived from the equations of propagative distorsion. The horn was divided into a number of cascaded segments, with either exponential or conical expansions ; the relative delay of the successive signal points was computed inside each segment and then summed to obtain the global acoustic delay. (Figure 6)

III. Implementation in a DSP56001

The next step was then to realize the computation in real time. The process could be implemented in two different manners :

- In time domain where each sample will be delayed or put forward in function of its position and its initial velocity (propagative correction), with a correction of its amplitude (static correction)
- Directly in the frequency domain, where the notion of harmonics is more natural, but which demands a larger power of calculation for achieving a FFT and IFFT

Mainly for reasons of hardware we have chosen to treat only the time domain.

III.1. Set up

Formulas of $\tau(\xi_0, l, \chi_0)$ and static correction were decomposed in power series, with a truncation which was determined by experimentation.

We have so two sets of coefficients : one for the static attenuation, and one for the delays calculation (this latter being the sum of the coefficients relatives to each geometrical parts of the horn taking in account the velocity at the beginning of each section).

This coefficient are the keys of the software, and must be carefully chosen. They will change with the modification of the conditions of exploitation (gain of the amplifier, initial acoustic velocity, geometry of the horn...) but otherwise are considered as constant.

III.2. Processing the data (figure 3)

After the sampling stage, each data is multiplied by the coefficients of the static power series adjusting its amplitude in function of its velocity.

After that stage the delay for each sample is calculated.

Delay implementation is performed using a circular buffer. As this process must be able to advance or delay the sample a general offset is done (half the length of the buffer). A sample may be delayed by steps of 1/fs (about 20.8 µs for fs = 48 kHz). Maximum delay is half the length of the buffer multiplied by the sampling period.

Each cycle the buffer rotate, output his leftmost sample and input the newest sample. This sample is inserted in the buffer corresponding to its delay, and the samples to its left are shift. When a delay is asynchronous to the clock rate (i.e. not a multiple of the sampling period), the corresponding sample is split between the adjacent samples in the buffer (« interpolation process »). A second parallel buffer (same length) store the weight of each sample.

Each output sample is divided by his weight and driven to the D/A converter.

IV. Experimentation (figure 2)

Equipment : FFT ONO SOKKI, Audio precision, An amplifier powerful enough to drive at high sound intensity a cabinet with the TAD driver loaded with a NEXO horn. A unique digital source was split into two parts to allow a direct AB comparison. Measurement and listening tests have been made with satisfactory results. The last results (sine wave burst signals) have shown an average diminution of distortion of 15dB in a bandwidth from 2kHz to 8kHz (with peak at -23dB near 3.5kHz) (figures 8 & 9). The listening test was less satisfactory as the unit wasn't driven as loud as we wanted to (and so have a lesser effect on the correction), but on some high pitched vocal a sensible difference have been made.

However at this point of development those tests were essentially needed to valid the model in a real time process.

IV.1.Propagative distortion correction

We have clearly observed that propagative distortion was the main cause of distortion in those tests. Main factor are :

- LEVEL :The reduction is proportionnal with the level, and best results have been achieved with sound pressure close to 120 dB. As seen before, the buffer that orders the samples has a certain time accuracy. If there is not enough pressure to have delays greater than that accuracy the correction will be less efficient. Figure 10 shows clearly that the THD+N is proportionnal to the level in a non-corrected signal, but on a corrected signal we noticed a threshold under correction is unactive and then the level of THD increase at a slower rate than the level of the signal (figure11).
- SAMPLING RATE : the sampling rate will determine the smaller correction possible (buffer accuracy). Figure 12 shows the difference between a sampling rate of 48 kHz and 80 KHz.
- PHASE : The correction process is highly dependant of phase. As we compute delays, a phase inversion would do the opposite of the correction and worsen the distortion ! (figure 13)

V. Further developments.

V.1. Model improvement

In the modeling we have considered that the medium was perfect and adiabatic, and the acoustic wave was plane. Such assumptions are quite close to the reality in most of the cases, however there is a certain point where approximations begin to reach their limit of validity. At this point the deformation of the wave (according to our model) if so strong that from a certain point the signal become non causal (figure 7). In this case the point of validity was fixed by two factors : initial velocity and frequency.

From this point a second order approximation may be not efficient enough, and the we should look after equation dealing with strong non linearities or dissipative medium [10].

The pain of such equations seems to be quite useless as the process seems to work without problems until 10 kHz. (That means that harmonics are out of the audio range, and 2^{nd} is hardly audible).

In conclusion we can think that this model works quite well for an audio application, and that the increase of performance are likely to be done with the implementation.

V.2. Implementation improvements

- In order to implement our algorithms on a limited DSP (56001 running at 20MHz), some trade off have been made : troncation of the power developped equations at low order, a process of time interpolation not so easy (and very heavy for the DSP as it concern division).
- As you have noticed, the way of shifting a sample in a buffer is close to the structure of an FIR. That could also explain why the correction is not as efficient on the very high band : the algorithm of correction sometime act like a low pass filter with a cut off frequency around 10 kHz(with only two or three taps, but the effect is noticeable). And as the coefficient have been computed for a certain velocity, the correction effect is too high (the signal is attenuated by this pseudo-filter) and consequenly have the inverse effect !
- In the state of the project, where distortion above 8 kHz is not corrected, and even slighly increased, a filtering could be made to split the high frequency band in 2 part. A corrected part (1-10 kHz) and a non corrected part (above 10 kHz) which would have a constant delay (to match the offset of the corrected part).
- As we seen before as the samling rate is very important an over sampled signal could be considered as being a major improvement.

VI. Conclusion

A new model of non linear behaviour in horns and phase plugs have been determined and succefully tested. If the theoretical model has some limits, those are well above the limit imposed by the technology which was used. Further test and developpement are planned that should confirm this point. The model and the process of precorrection have been patented in several countries.

VII. References

[1] BERANEK « Acoustics » McGraw-Hill, New York 1954

[2] H.F. OLSON « Acoustical Engineering » D Van Nostrand1957

[3] A. JINDAL « Simulation of nonlinear propagation of finite amplitude sound wave through a circular pipe. » Acustica Vol.38 1977

[4] Philippe ROBINEAU « Device for processing an audio frequency electrical signal » NEXO patent N°86.16.244 - 1986

[5] Julian DUNN « Prototyping CDA with a DSP56001 » Prism sound Limited, internal document. - 1992

[6] Rémi VAUCHER « Implementation d'une correction de distorsion dans un DSP56001 » -ENSAM thesis 1993

[7] SCHMITT Regina ; KLIPPEL Wolfgang « Modeling of the nonlinear Behaviour of a Horn Loaded Compression Driver System » AES preprint n°3256

[8] SCHURER Hans ; BERKHOFF A.P. ; SLUMPC.H. ; HERRMANN O.E « Modelling and compensation of non-linear Distorsion in Horn Loudspeaker » AES Preprint n°3819

[9] Y. ROCARD « Dynamique générale des Vibrations » ed. Masson 1960

[10] B. POIREE - D. ODERO « Revue du Cethedec N°46 » 1976

VIII.Figures

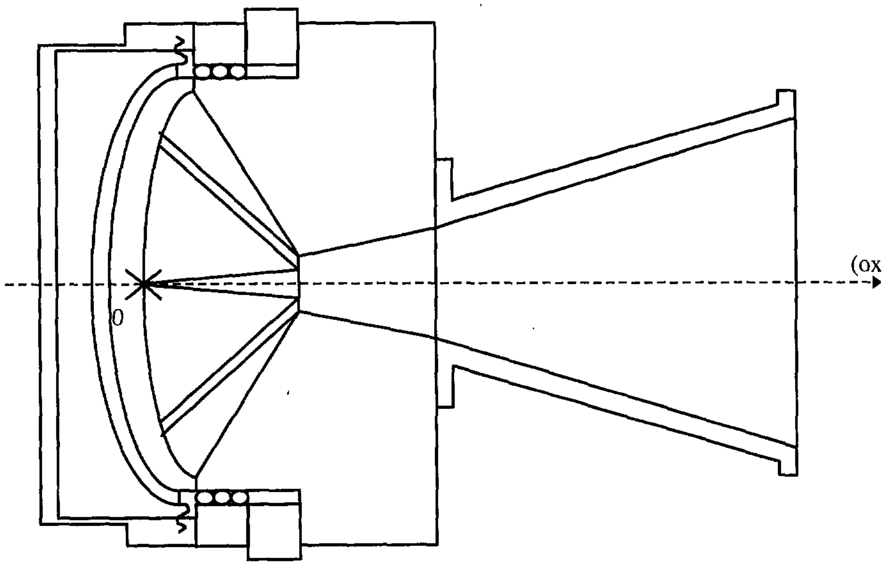


Figure 1

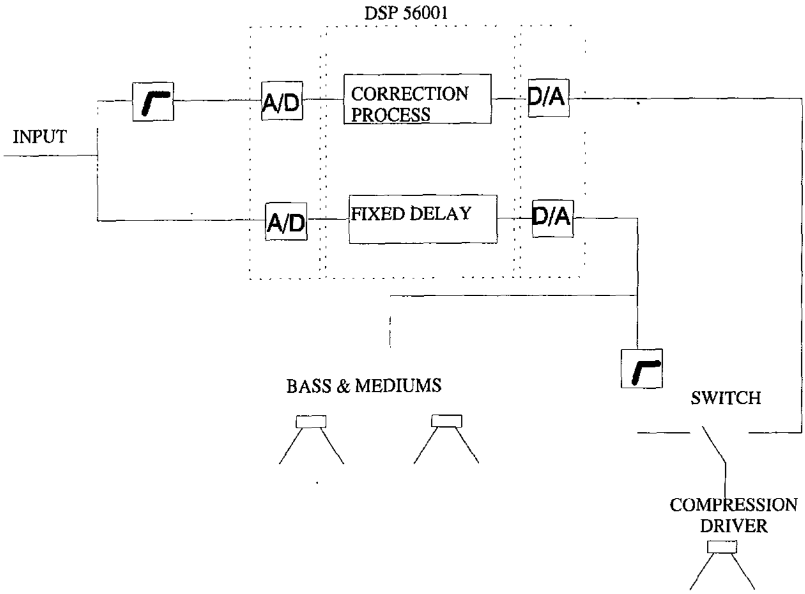

DSP 56001
INPUT
A/D
CORRECTION
PROCESS
D/A
A/D
FIXED DELAY
D/A
BASS & MEDIUMS
SWITCH
COMPRESSION
DRIVER

Figure 2

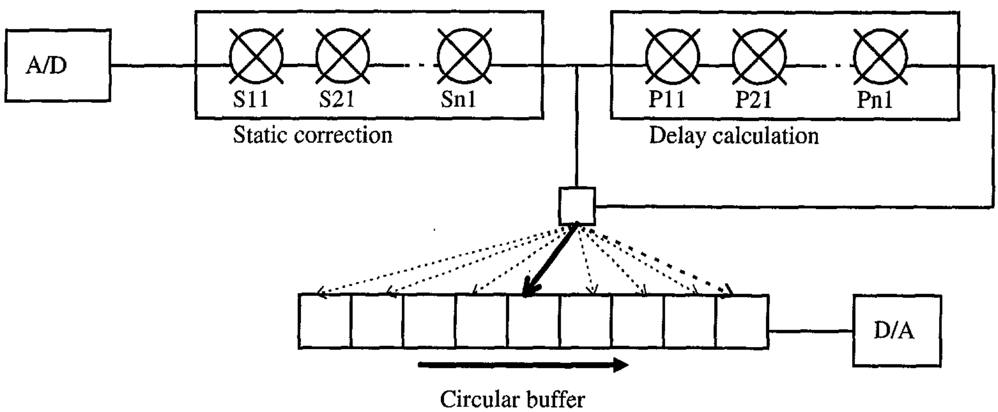

A/D
S11
S21
Sn1
Static correction
P11
P21
Pn1
Delay calculation
D/A
Circular buffer

Figure 3

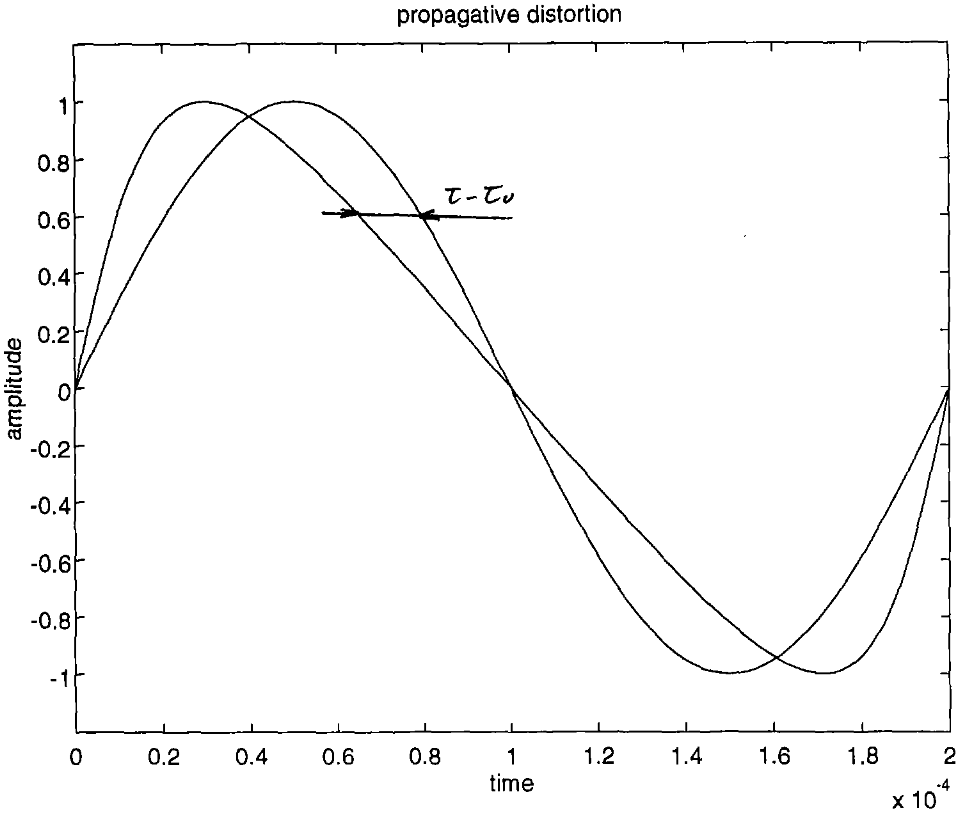


Fig 4

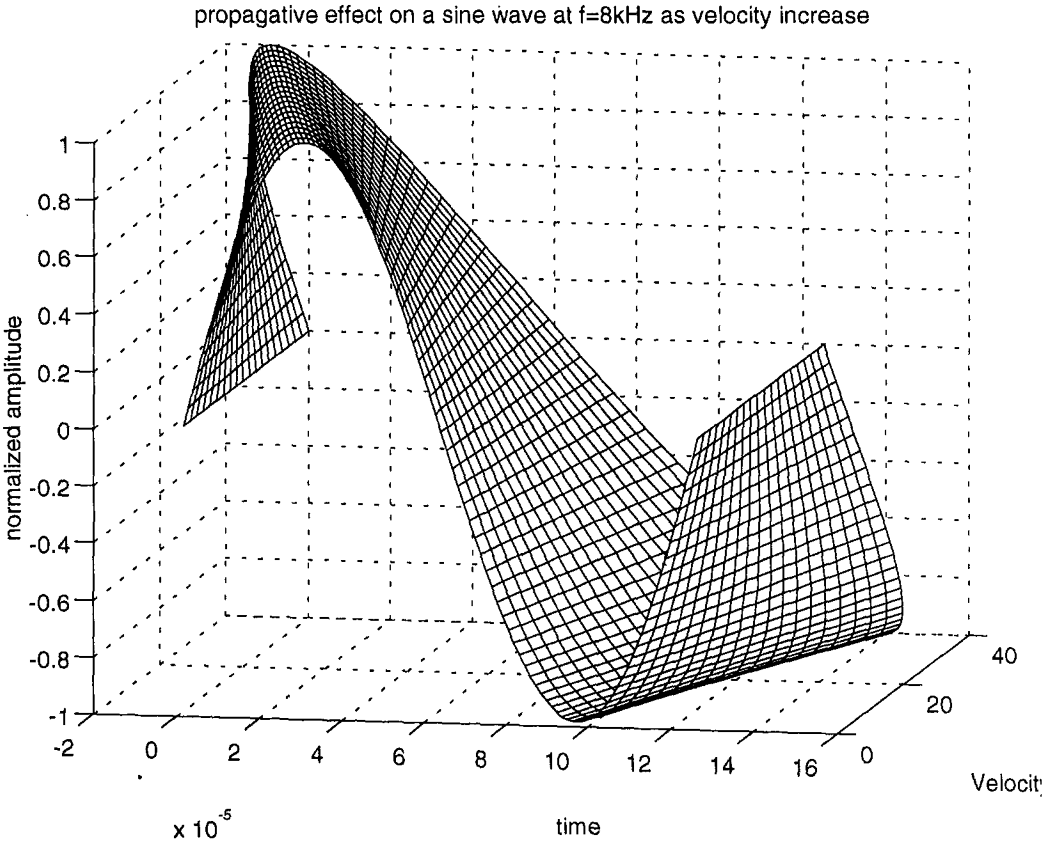

propagative effect on a sine wave at f=8kHz as velocity increase
normalized amplitude
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-2
0
2
4
6
8
10
12
14
16
x 10^-5
time
0
20
40
Velocit

Fig. 5

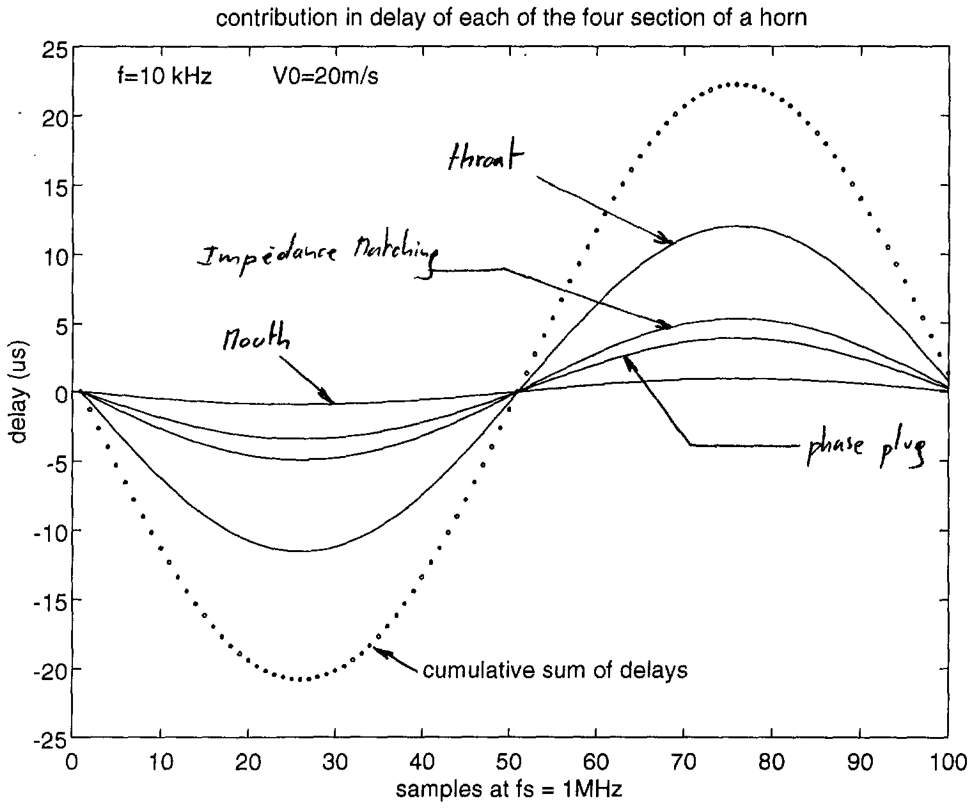

contribution in delay of each of the four section of a horn
f=10 kHz
V0=20m/s
throat
Impedance Matching
Mouth
phase plug
cumulative sum of delays
delay (us)
samples at fs = 1MHz
25
20
15
10
5
0
-5
-10
-15
-20
-25
0
10
20
30
40
50
60
70
80
90
100

Fig 6

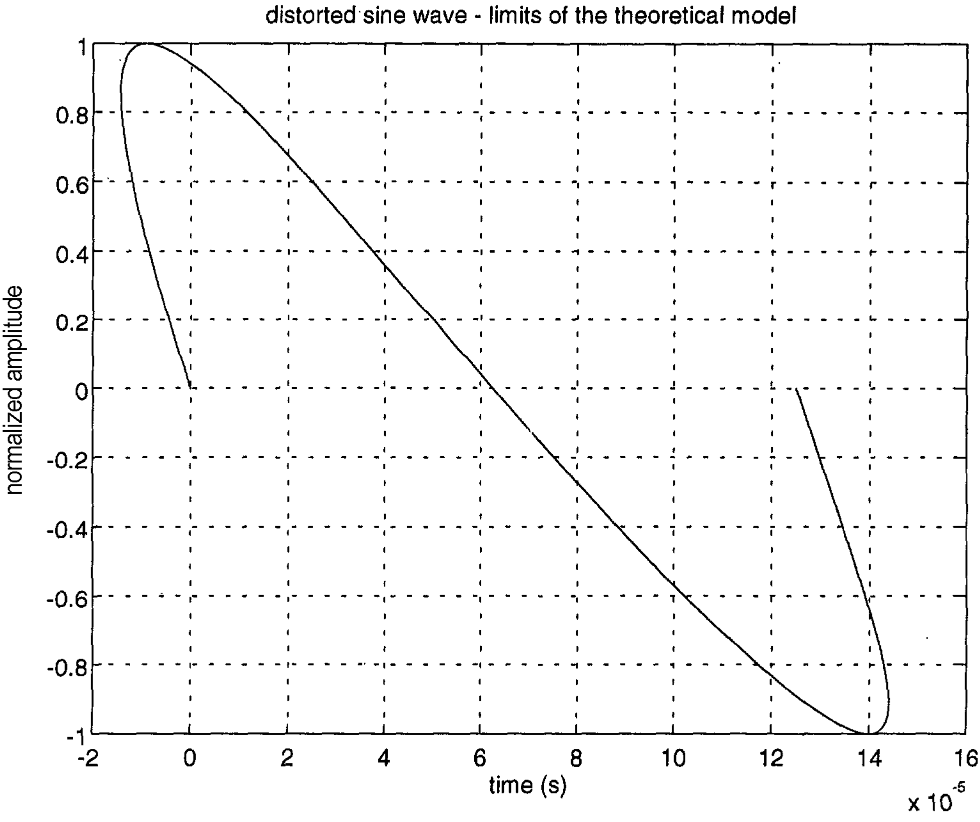
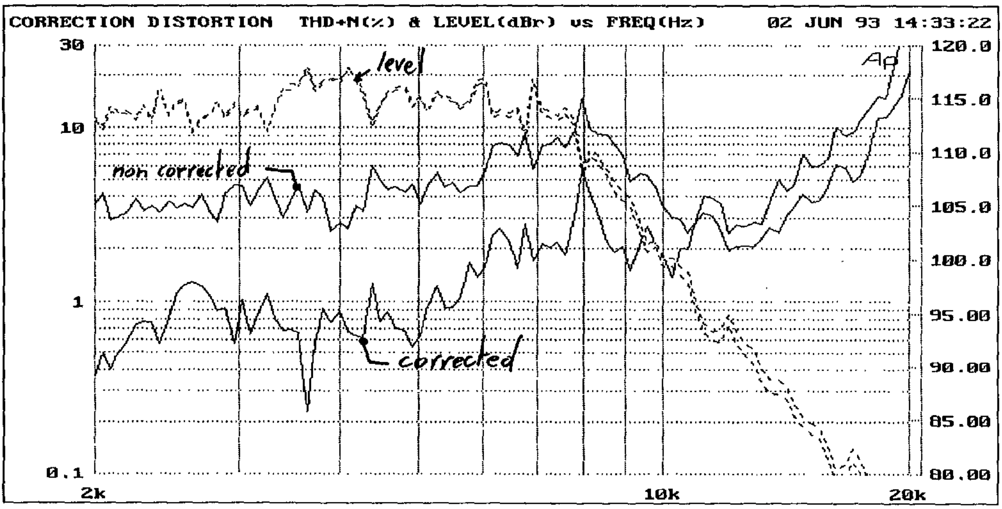
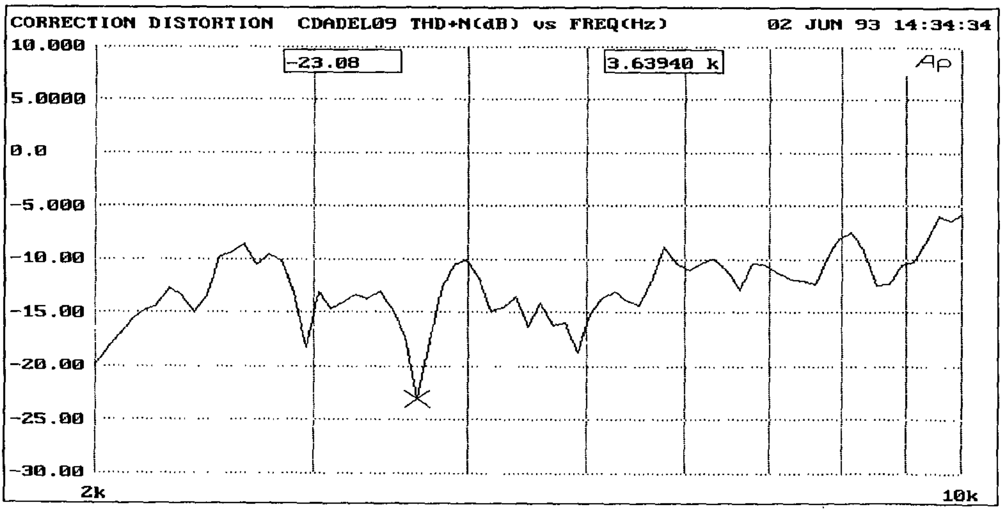

distorted sine wave - limits of the theoretical model
normalized amplitude
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-2
0
2
4
6
8
10
12
14
16
time (s)
x 10^-5

Fig. 7



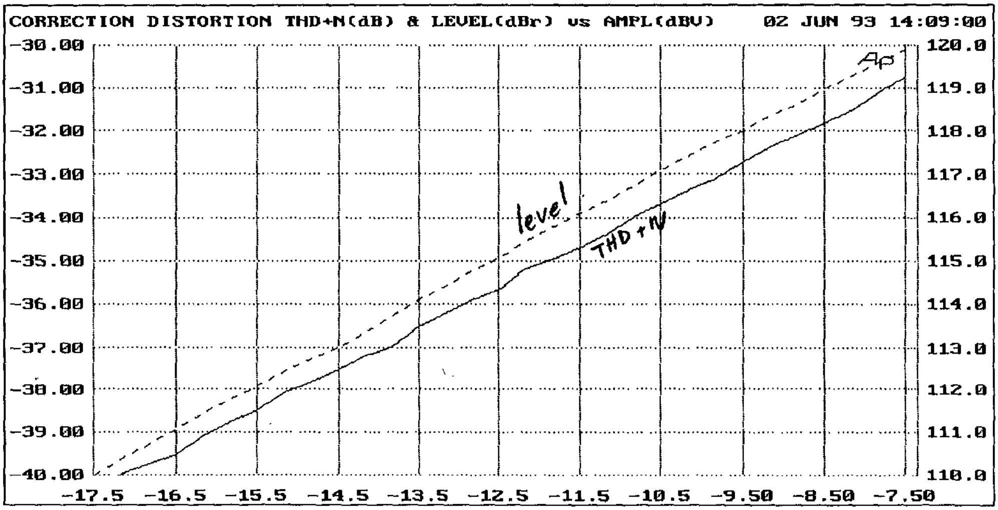
NEXO MSIC2 + TAD driver
High pass filter before the process
corrected and non corrected signals

Fig 8



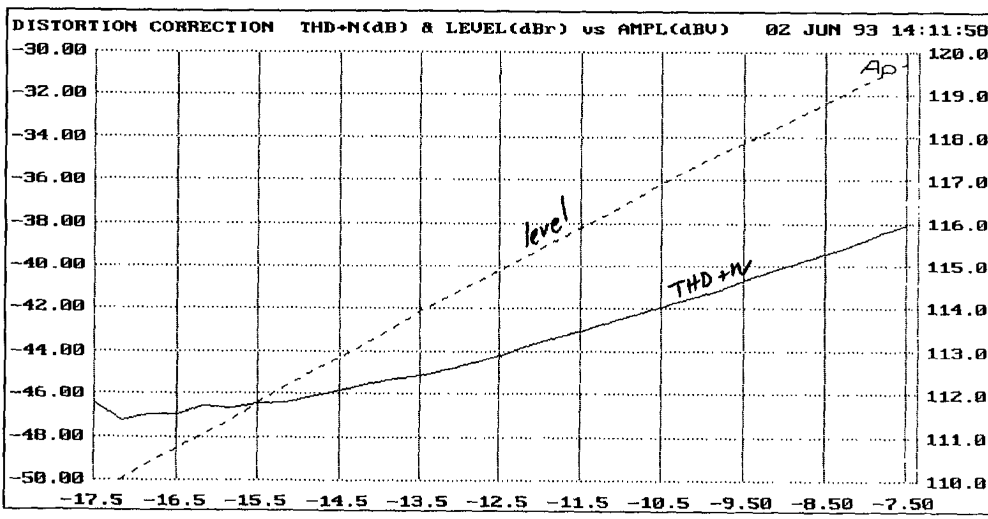
NEXO MSIC2 + TAD driver
High pass filter before the process
Difference between the corrected and uncorrected signals

Fig 9



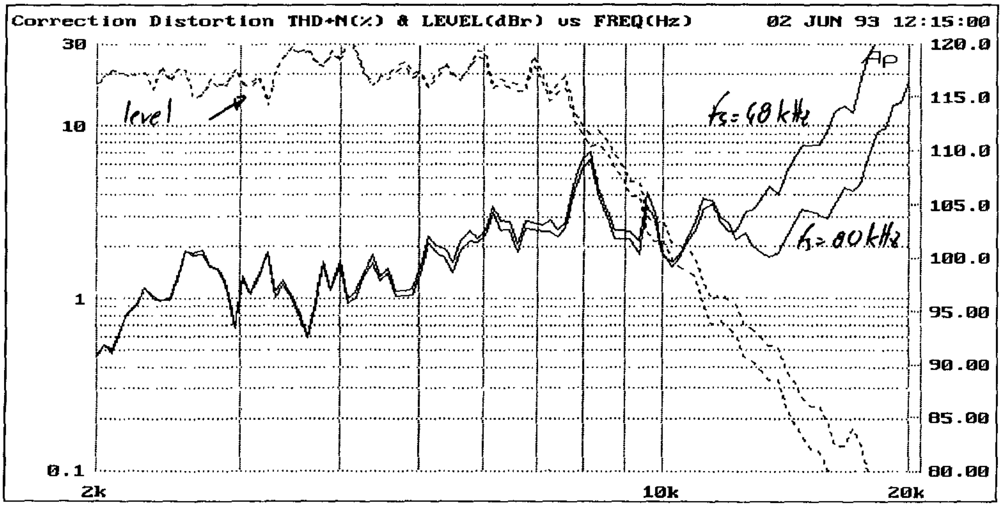
NEXO MSIC2+TAD driver
non corrected signal, linear progression of THD versus level.

Fig 10



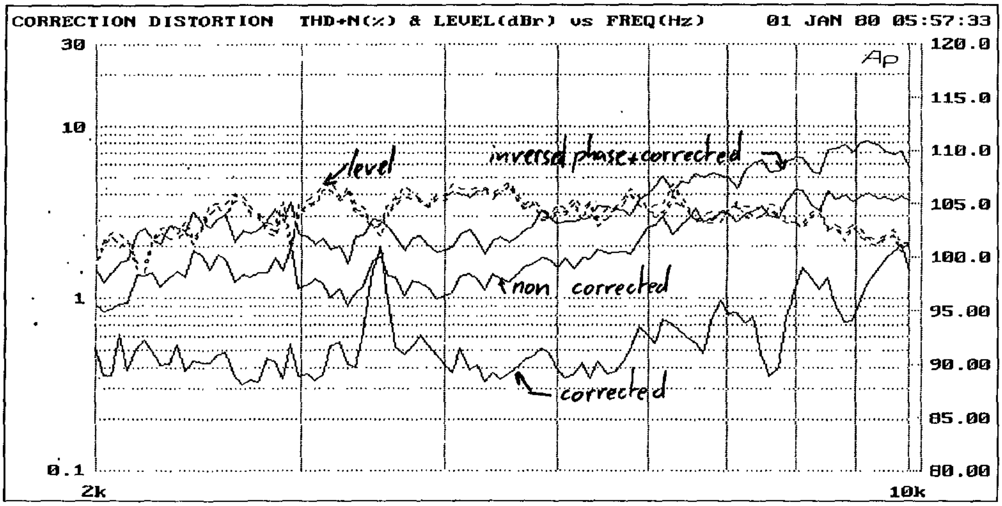
NEXO MSIC2 + TAD driver
corrected signal, THD versus Level

Fig. 11



NEXO MSIC2 + TAD motor
Low pass filter 8 kHz before the process
corrected signal at fs=48 kHz and 80 kHz

Fig. 12



NEXO MSIC2 + TAD driver
Middle signal is uncorrected signal
Upper signal is the corrected signal with an inverted phase
Lower signal is the corrected signal

Fig. 13