POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Compensation of phase distortion in high-performance audio systems

Laurea Magistrale in Music and Acoustics Engineering - Ingegneria della Musica e dell'Acustica

Author: Martino Schgor

Advisor: Prof. Giuseppe Bertuccio

Co-advisor: Marco Tagliaverga

Academic year: 2020-2021

Introduction

Hi-Fi audio systems are usually designed with extreme care, in order to reproduce the recorded sound information as accurately as possible, with extremely strict tolerances. The pursuit of high fidelity has lead audio engineers over the years to face all the non-idealities in the transfer function of audio systems. However, the information about the phase response is hardly taken into consideration, thought by many to have no audible effect. This work addresses the phenomenon of phase distortion by highlighting its causes and its perceptual effects on audio quality. Eventually, a close-loop approach at the compensation of phase distortion is proposed, with the use of a digital signal processor.

1. Formal definitions

Linear Time-Invariant transfer functions are determined by their magnitude and phase responses in frequency domain. They may have a nonzero phase response for several reasons: a pure delay, for example, introduces a phase term proportional to frequency. Other kinds of phase changes might be related to a frequency filtering (the minimum-phase components, that can be evaluated through the Hilbert relation), an all-pass component, or a polarity reversal.

It is necessary to tell apart the *distortionless* characteristics (such as a pure time delay) from the ones that are related to an actual modification of the signal waveform. The combination of the latter is usually called *phase distortion*.

The phase response of a distortionless system is proportional to the frequency or *linear*. Suitable all-pass filters can be designed with the intent of linearizing the phase of a system.

Several measures are proposed in literature to quantify the effect of phase distortion, the absolute peak of the phase response, as well as *phase delay* and *group delay* are the most commonly used.

2. Causes

Near all modern audio systems work with an amplifying circuit and electrodynamic loudspeakers acoustically loaded by their enclosures. While the electronic amplifiers have little phase effects on the signal, the mechanical behaviour of the loudspeaker, the presence of filters and the acoustic loading may introduce an appreciable phase distortion. Moreover, ported enclosures, passive radiators or multi-driver systems have to take account of the interference pattern gen-

erated by the use of multiple sound field sources. The same phenomenon of interference happens naturally between the sources of a stereo system, or considering the room reverberations, but these effects might be appreciated for the generation of a natural perceptual *soundstage*.
Most of the behaviour of a system can be predicted with a simulation method, the aleatory part of the model can be reduced effectively with the use of a multi-way system and crossover filters.

3. Audibility

The audibility of phase distortion has been a topic of discussion for decades. The most recent studies agree that there are some effects of phase audibility, however, they are way less impacting the sound quality with respect to other nonidealities, like nonlinear distortion or frequency filtering. In particular, monaural phase effects are said to have a small impact on sound timbre even for steady-state sounds, while binaural phase is used in the process of sound source localization.

3.1. Monaural auditory models

Many models have been proposed in scientific literature, to formalize the process of hearing. One of the most famous is the *filterbank model*, that offers the advantages of being coherent with the human physiology and effective for perceptual audio coders. Such model cannot explain any phenomenon of phase audibility.
More sophisticated models involve some degree of phase audibility, the *in-band correlators* proposed by Licklider [4] seem to make the simplest explanation.

3.2. Binaural auditory models

Given the high capability of humans of identifying the direction of arrival of a soundwave, the mechanisms involving binaural phase comparison cannot be neglected. A first model was proposed by Jeffress [3] consistent with the observation of neural topology in birds. However, the same neural disposition is not observed in human beings. A more empirical model has been proposed by Biberger [1] to justify some phenomena of binaural unmasking.

Figure 1: Frequency response near-field measurement of a TMAUDIO R2c system

4. Preliminary experiments

4.1. Loudspeaker system build

A simple loudspeaker system was developed and built with all the optimized characteristics for studying loudspeaker phase distortion. It featured single-way drivers in sealed box, to avoid all the possible effects of crossover filtering, acoustic remixing in crossover band, port phase shift and port interference. This activity was part of the project TMAUDIO R2c, developed at TagMa S.r.l.s in Milan, Italy.
Its linear behaviour was both simulated and measured in magnitude and phase response, showing the difference between the predictable and aleatory phenomena. The unavoidable phase distortion was found and measured. Figure 1 shows the setup for the frequency response measurement, while the results of both the simulation and measure are shown in figure 2.

4.2. Audibility Experiences

A few informal perceptual experiments about monaural phase audibility have been recreated, with the aim of validating the auditory mod-

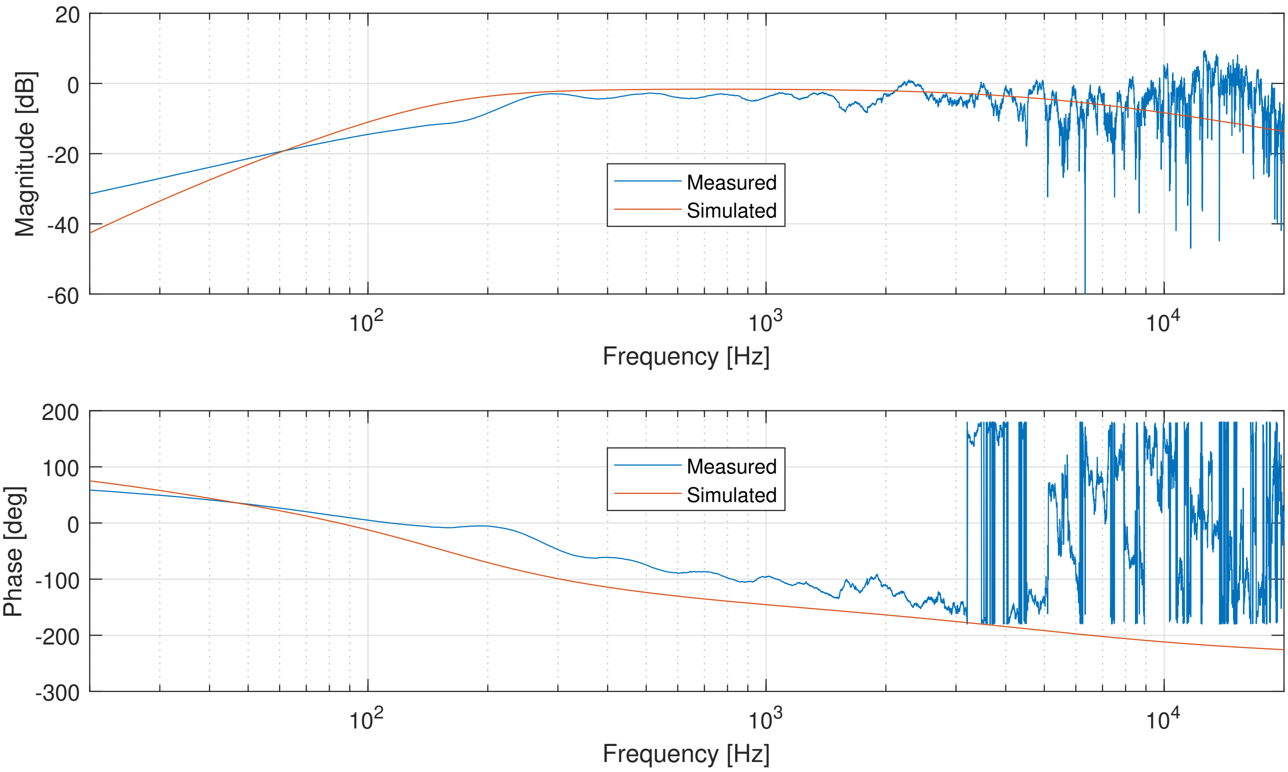

Figure 2: Simulated and measured frequency response of TMAUDIO R2c. The measured phase response is kept in wrapped form for graphical reasons.

els. The signals described by Miller and Taylor, Schouten and Patterson [5–7] have been generated in MATLAB and listened by the author and a few other volunteers. The evidence for monaural phase effects is extremely clear for those signals.

4.3. Audibility Measurement

We attempted to perform a formal measurement of phase audibility for specific signals, in order to identify the most suitable measure for phase audibility and the average thresholds taken from a large number of listeners. Due to the Covid 19 pandemic, the experiment was carried out through an online web app, with no control on the playout system. The formal experiment was a failure due to this lack of control, but the data suggested that there might be a learning process in the detection of phase effects. The author himself performed many listening test in the attempt to set an audibility threshold, but the data converged to a curve after hundreds of tests.

5. Compensation

5.1. Models of transfer functions

There might be disagreement on the choice of transfer function on which we want to force linear phase response, so here are proposed four different models of correction:

1. **Headphones** When listening to headphones, the sound pressure at the ears coincides with the one in proximity of the drivers, as no appreciable channel mixing occurs. We will then consider the transfer function from the electrical analog signal to the sound pressure inside the pavillions and attempt to linearize its phase response.
2. **Loudspeakers** The second model aims at the correction of loudspeaker non-idealities only. The chosen transfer function is from the analog signal to the sound pressure generated at the mouth of the loudspeakers, as close as possible to the drivers and on their axis. To avoid confusion with multiple sound field sources, a pair of single-way sealed-box loudspeaker systems will be used, the TMAUDIO R2c.

3. **Loudspeakers and room acoustics** The room reverberations cause the listener to hear a mix of the original audio signal and delayed and dimmed instances of it. Thus, their effect is often modelled as a convolutional filter, showing an appreciable frequency filtering and a nonlinear phase response. This time we pick the transfer function from each loudspeaker to the approximate position of the listener's head. The channel mixing happening at the ears is not considered in the transfer function.
4. **Acoustic unmixing** This last model addresses the acoustic channel mixing that occurs in space when both the sound sources are made to play simultaneously. The interference phenomenon is responsible for a different transfer function at each point in the space. Since we are only interested in the points where the ears are located, we will choose the two transfer functions from the stereo audio channels to the sound pressure directly at the listener's ears. In this case, we both need to linearize the phase response and to perform a magnitude equalization.

5.2. Hardware Setup

The DSP used is a Bela Board, designed by the English company Augmented Instruments Ltd. A preliminary testing of its preamplifiers and output circuits has been performed to ensure reliability of these stages, in terms of spectral flatness and linear phase over the audible range. Two sets of stereo microphones have been built out of CMA-6542TF omnidirectional electret capsules. Each capsule has been compared with a professional measurement microphone (OmniMic V2) for calibration purposes. One set features the capsules directly soldered at the extremities of a long coaxial cable, while the other consists in a wearable headset, to help positioning the transducers as close as possible to the ears.

5.3. Software

The block scheme in figure 3 shows the simplified architecture of the main software. Modifications are introduced for optimization purposes or to adapt to the first three out of the four models. Every 186 ms approximately, a 8192-sample window is fetched from the internal memory and transformed in frequency domain, then phase-shifted using a stored estimate of the loudspeaker-to-microphone transfer function, retransformed in time domain and played out. Finally, the transfer function is updated by comparing the output signal with the recording from the microphone.
This implementation works for steady-state signals, it has been tested with a 440 Hz square wave and a visual evaluation on the oscilloscope. However, the phase shift is instantaneous at the margin of each window, generating an annoying clicking effect. The processing scheme has been modified to implement an *Overlap and Add* (OLA) technique with Bartlett windows to overcome this issue.
The full block diagram of the OLA version is reported in figure 4.

5.4. FFT-powered compensation

The perceptual experiment was conducted by asking 12 volunteers (all with normal hearing and most well trained to active listening) to participate in the following listening experiences:

1. **Headphones** The microphone headset was used, in combination with a pair of OMEN hp 800 headphones. The listeners were given control of a GUI, with the possibility of bypassing the correction mechanism. Some musical pieces were played through the system.
2. **Loudspeakers** The experimental setup was similar to the previous case, with the only difference that the microphones were placed in front of a pair of loudspeakers (TMAUDIO R2c).
3. **Loudspeakers and room acoustics** The GUI was modified to have 3 states: the first one related to calibration, in which wideband signals were played (one channel at a time) and the transfer functions updated. Then the correction was in open-loop, with the usual possibility of bypass mode.

5.5. XTC Attempt and Localization Technique

To correct the full transfer functions to the listener's ears, it is necessary to counteract the *crosstalk* i.e. the effect of each sound source at the contralateral ear. The scientific literature reports some techniques of *Crosstalk*

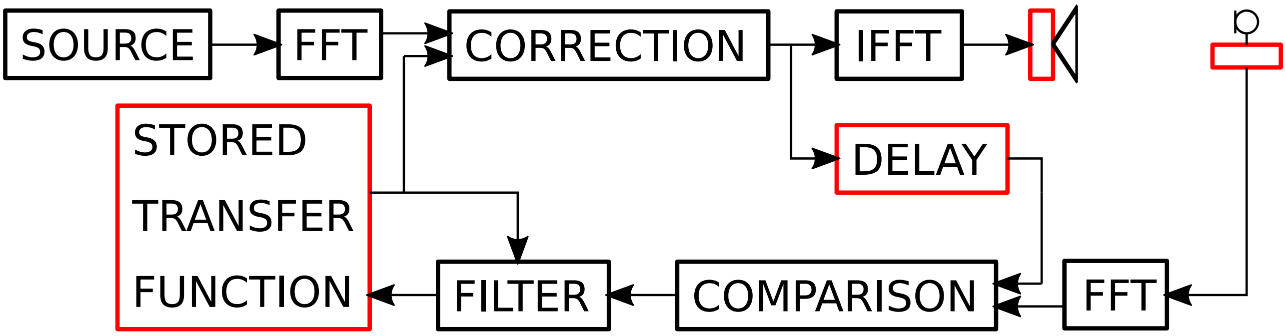

Figure 3: Simplified block diagram of the processing. The latched blocks have been represented in a red frame.

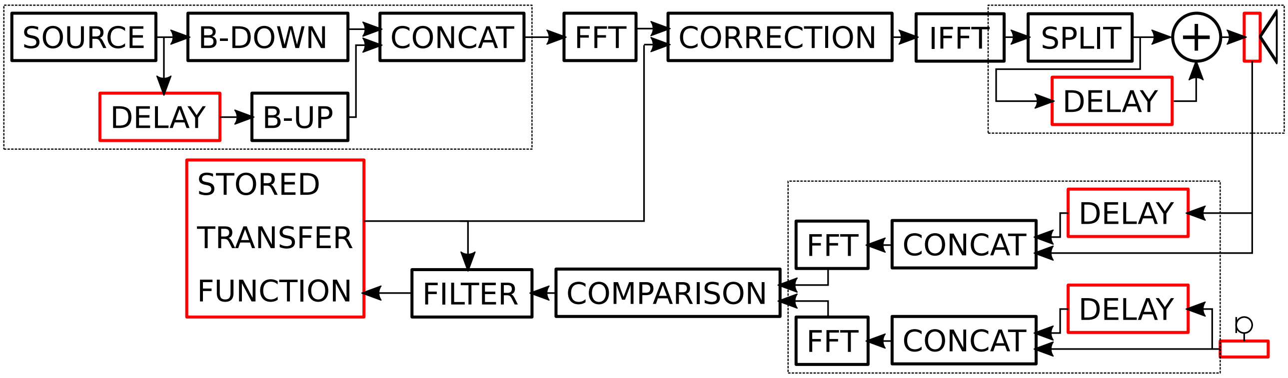

Figure 4: Diagram of the Overlap-and-Add version of the compensation software. The "B-UP" and "B-DOWN" blocks represent the multiplications with a linear functions, namely the upwards and downwards halves of a Bartlett (triangular) window.

Cancellation (XTC) that always require a precise localization of the ears, usually obtained with the use of optical sensors.
An attempt was performed with the technique from Choueiri [2], slightly modified so that the optical localization is not needed anymore. In fact, we propose a microphone localization technique based on *Generalized Cross-Correlation* (GCC) adapted to work in a listening room with a stereo system.
The GCC techniques found in literature are mostly used for source localization with an array of microphones, and require the signals from distinct sources to be uncorrelated. It is not the case for a stereo system, where the channel correlation is usually high for musical signals. The idea is to consider the common and differential modes of the channels, and decompose each of them into a component that is correlated with the other and an uncorrelated residue. By performing GCC between the microphone input and the residues, we obtain an estimate of the common and differential modes of the transfer functions from the speakers to the microphone. More detailed information can be found in the thesis.

5.6. Results

We report the results from the four listening experiments

1. **Headphones** All the listeners stated that the phase correction had a negative impact on quality, due to an unpleasant perception of "sound coming from inside the head".
2. **Loudspeakers** The effect of correction of loudspeaker systems is clearly more subtle than the headphones case. Some listeners did not hear any difference introduced by the DSP, while the majority of the others still preferred the bypass version.
3. **Loudspeaker and room acoustics** None of the listeners could tell apart the real-time closed-loop correction from the compensation based on the pre-recorded transfer function, so the results of this experi-

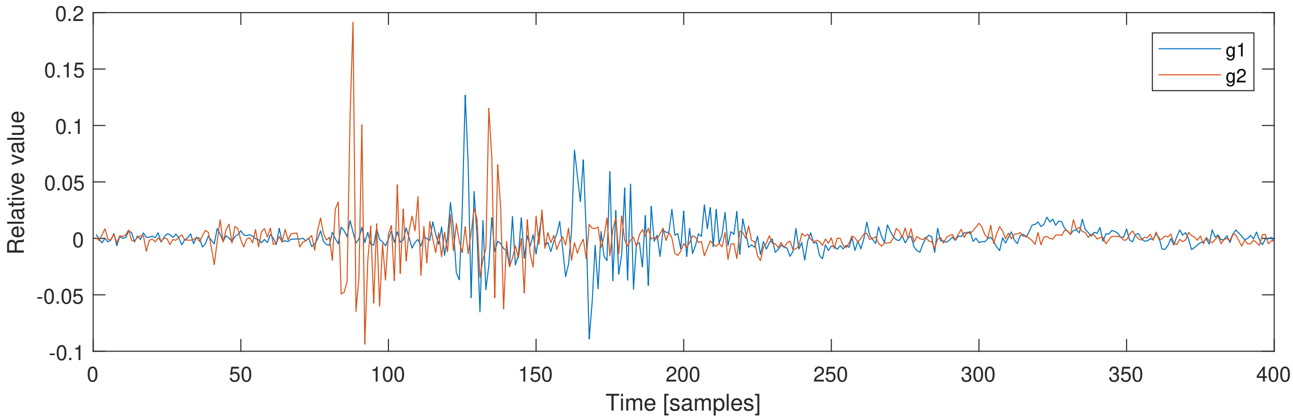

Figure 5: Estimated time-domain impulse responses (from two loudspeakers to a microphone) using the proposed localization technique. The main peaks are clearly detectable, and some secondary artifacts arise from the reverberations of the room.

ment coincide with the previous.

4. **Acoustic unmixing** The XTC technique did not work, probably due to the room reverberation. In fact, many experiments about XTC are performed in anechoic rooms only. The formal listening test was not carried out. However, this attempt led to the development of the microphone localization technique, that works with good reliability in reasonably reverberant rooms. An example of estimated impulse response using this technique is reported in figure 5.

Conclusions

This work highlighted that phase distortion happens in almost every kind of audio systems, even the most expensive. Its perceptual effects are way more subtle with respect to other waveform distortions, such as nonlinear distortion or frequency filtering, nevertheless, we can state that monaural phase is audible, and a reasonable psychoacoustic model can give an explanation to this phenomenon. The measuring for monaural phase perception is hard to perform, the measures that were proposed in scientific literature cannot be considered absolute and the perception requires a certain degree of attention from the listener.

Binaural differential phase distortion is more easily perceivable, not related to a timbric change, but rather responsible for the generation of the soundstage. However, negative effects on the differential phase can be easily avoided by respecting the symmetry of the audio system build.

The last experiment was a failure because the setup was extremely sensitive to model non-idealities. The experience should be performed in a quiet anechoic room but since we had no access to such an environment, the experiment is left for later developments. However, it might be worthy to notice that the microphone localization technique that has been developed in the preparation of this setup works with high reliability.

The first compensation experiments, have been carried out with the most interesting results. The correct functioning of the setup was guaranteed by the virtual oscilloscope, but the overall listening experience was considered worse by the majority of the listeners who declared themselves able to spot the difference. A possible reason could be that humans are so used to listening to a certain pattern of phase distortion that the compensated version may sound unnatural. In any case, further research will be needed to find a plausible explanation. We cannot completely reject the idea that phase distortion should not be regarded as a dangerous non-ideality, but rather as a parameter that can be artfully mastered by electroacoustic engineers. Such statement clashes with the obsessive search for "fidelity" often shown by audiophiles, but there might be a point where the pursuit of perfection gives way to the more meaningful mastery of the good sounding imperfections.

References

[1] T Biberger and S.D. Ewert. *Towards a Generalized Monaural and Binaural Auditory Model for Psychoacoustics and Speech Intelligibility.* Acta Acustica, 06 2021.

[2] E. Y. Choueiri. Optimal crosstalk cancellation for binaural audio with two loudspeakers. Self-published, URL `http://www.princeton.edu/3D3A/Publications/Choueiri_3D3A_OptimalXTC.html`, 2010.

[3] L.A. Jeffress. A place theory of sound localization. *J. Comp. Physiol. Psychol*, (41):35–39, 1948.

[4] J.C.R. Licklider. A duplex theory of pitch perception. *Experientia*, 7:128–134, 1951.

[5] George A. Miller and Walter G. Taylor. The perception of repeated bursts of noise. *The Journal of the Acoustical Society of America*, 20(2):171–182, 1948.

[6] R.D. Patterson, K. Robinson, J.L. Holdsworth, D.A. McKeown, Zhang C.Q., and Allerhand M. Complex sounds and auditory images. Published in: Auditory physiology and perception, Proc. 9th International Symposium on Hearing, Eds: Y Cazals, L. Demany, and K. Horner. Pergamon, Oxford, 429-446., 1992.

[7] J. F. Schouten, R. J. Ritsma, and B. Lopes Cardozo. Pitch of the residue. *The Journal of the Acoustical Society of America*, 34(9B):1418–1424, 1962.