Virtual Bass Enhancement via Music Demixing

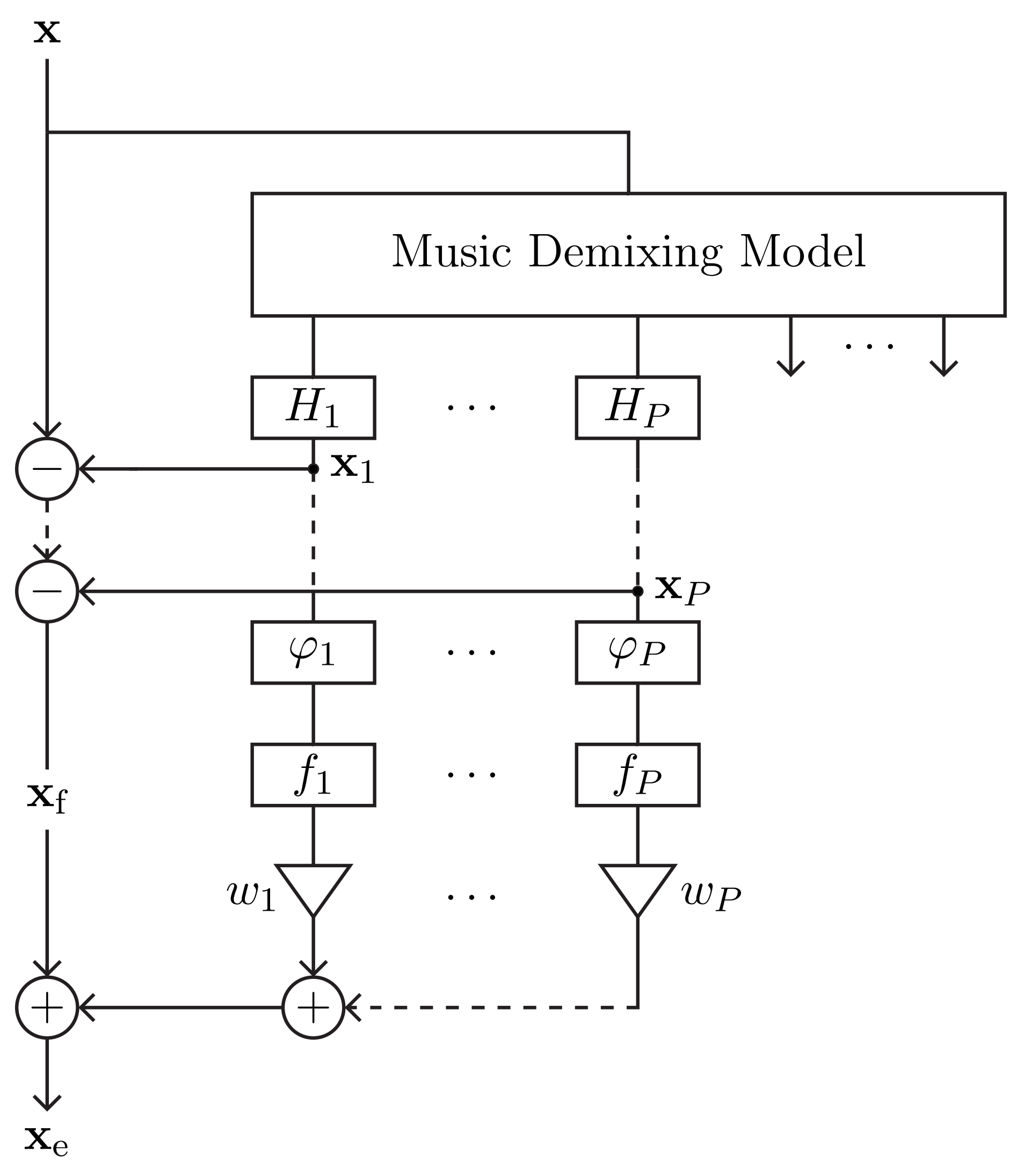
Riccardo Giampiccolo, Alessandro Ilic Mezza*, Alberto Bernardini, and Augusto Sarti*
Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, Italy
{riccardo.giampiccolo, alessandroilic.mezza, alberto.bernardini, augusto.sarti}@polimi.it

1. Abstract

Virtual Bass Enhancement (VBE) refers to a class of digital signal processing algorithms that aim at enhancing the perception of low frequencies in audio applications. Such algorithms typically exploit well-known psychoacoustic effects and are particularly valuable for improving the performance of small-size transducers often found in consumer electronics. Though both time- and frequency-domain techniques have been proposed in the literature, none of them capitalizes on the latest achievements of deep learning as far as music processing is concerned. We propose a novel time-domain VBE algorithm that incorporates a deep neural network for music demixing as part of the processing pipeline. This technique is shown to improve the bass perception and reduce inharmonic distortion, i.e., the main issue of existing time-domain VBE algorithms. The results of a perceptual test are then presented, showing that the proposed method is able to outperform state-of-the-art algorithms both in terms of bass enhancement and basic audio quality.

2. Proposed Method (1)

The figure shows the general block diagram of the proposed VBE algorithm [1]. The target audio frame $\mathbf{x}$ is processed by a Music Demixing Model (MDM) which extracts S stems. Then, P stems containing low frequencies are filtered by means of $H_p(z)$ and subtracted from the input frame $\mathbf{x}$ and passed through two types of functions: normalization functions φ_p and Nonlinear Devices (NLDs) f_p. The processed stems are finally weighted by w_p and summed back to $\mathbf{x}_\mathrm{f}$, yielding the bass-enhanced audio frame $\mathbf{x}_\mathrm{e}$.

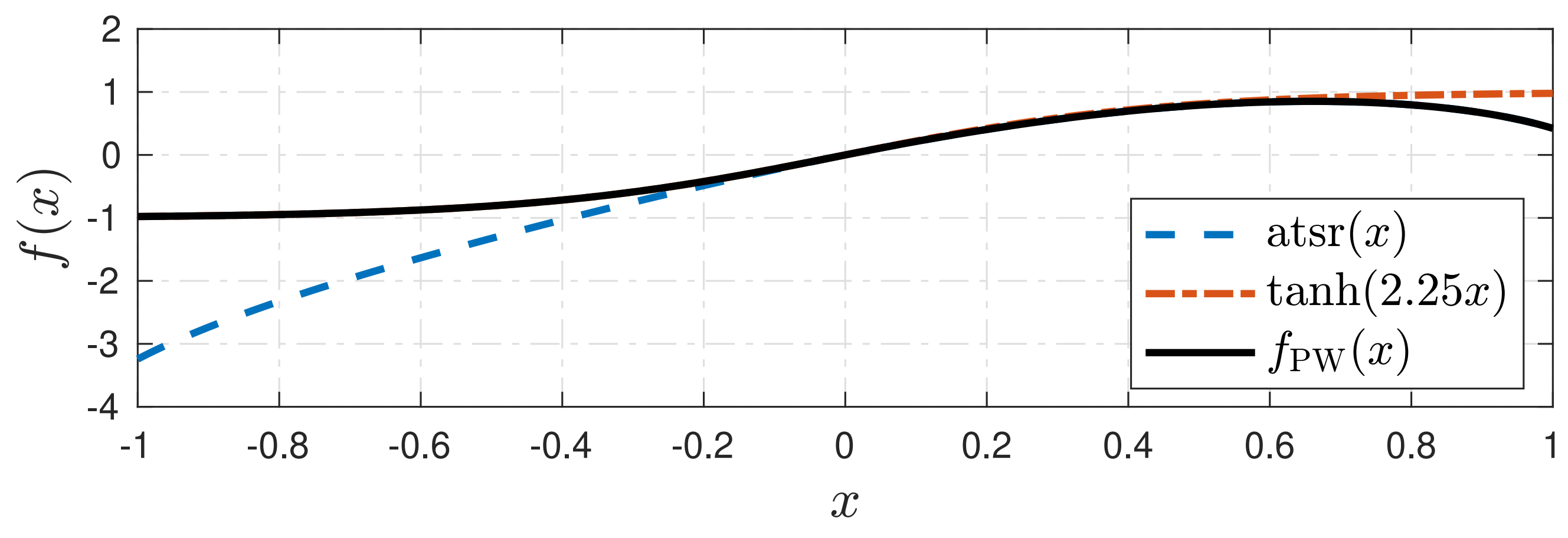


The overall processing pipeline can be written as follows:

$$\mathbf{x}_\mathrm{e} = \mathbf{x} + \sum_{p=1}^{P} \left(w_p \, \mathbf{f}_p(\varphi(\mathbf{x}_p)) - \mathbf{x}_p \right), \quad p = 1, \ldots, P\,.$$

Music Demixing Model
In our implementation, we select Spleeter by Deezer as MDM [2]. In particular, Spleeter consists of pre-trained 12-layer U-Nets (one for each stem) employed to estimate spectro-temporal soft masks suitable to separate single sources [2]. We extract and process just the `bass` and `drums` stems.

Filtering Stage
We employ a 10th order zero-phase forward-backward lowpass filter with cut-off frequency 250 Hz to extract the low end (kick drum and toms) from the `drums` stem. Zero-phase filters are considered to avoid phase distortion, which may impair the final result.



Proposed Nonlinear Device
NLDs are responsible for the harmonic generation that triggers the "missing fundametal" phenomenon. We propose a novel nonlinear device f_PW defined as

$$f_\mathrm{PW}(x) := \begin{cases} \mathrm{atsr}(x), & \text{if } x \geq 0 \\ \tanh(2.25x), & \text{if } x < 0 \end{cases},$$

with

$$\mathrm{atsr}(x) := 2.5 \tan^{-1}(0.9x) + 2.5\sqrt{1-(0.9x)^2} - 2.51\,.$$

Such an NLD combines in a piecewise fashion two known NLDs [3] with the aim of preventing loss of headroom while generating, at the same time, odd and even harmonics.

2. Proposed Method (2)

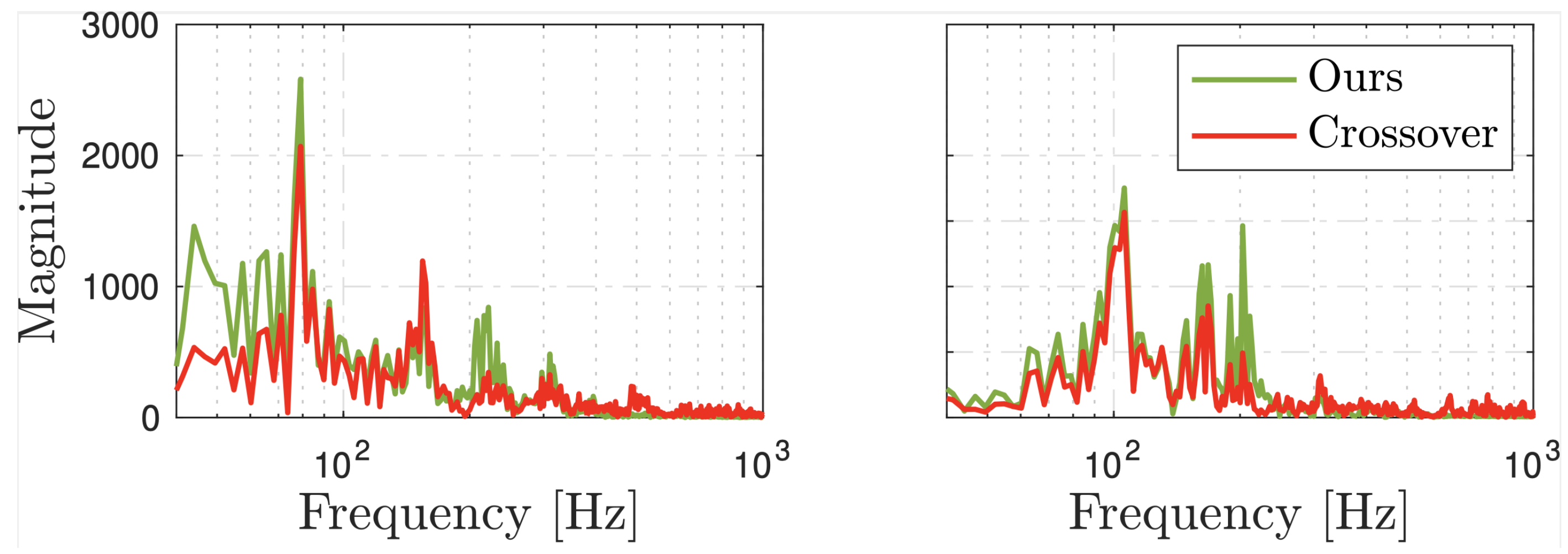
Normalization and Output Stage
Depending on the amplitude of the input signals, the NLDs risk being mostly visited in their quasi-linear regions, severely impairing the generation of harmonics [4]. To sort this issue out, we introduce the normalization functions

$$\varphi_p(\mathbf{x}_p) := \frac{\beta_p}{\max\left\{ |\mathbf{x}_p[j]| \right\}_{j=k}^{k+K-1} + \varepsilon} \, \mathbf{x}_p\,, \quad p = 1, \ldots, P\,,$$

where $\beta_p \in (0, 1]$.
Finally, we can write the enhanced audio frame $\mathbf{x}_\mathrm{e}$ referred to our implementation as

$$\mathbf{x}_\mathrm{e} = \mathbf{x} - \mathbf{x}_1 - \mathbf{x}_2 + w_1\, \mathbf{f}_1(\varphi(\mathbf{x}_1)) + w_2\, \mathbf{f}_2(\varphi(\mathbf{x}_2))\,.$$

The figure shows the Fast Fourier Transform (FFT) magnitude of the difference between $\mathbf{x}_\mathrm{e}$ and $\mathbf{x}$ for two audio tracks considered in the perceptual test. With respect to the crossover network [3] (one of the baselines), our method generates higher amplitude harmonics in the mid-frequency range of the spectrum, i.e., those responsible for the "missing fundamental" effect.



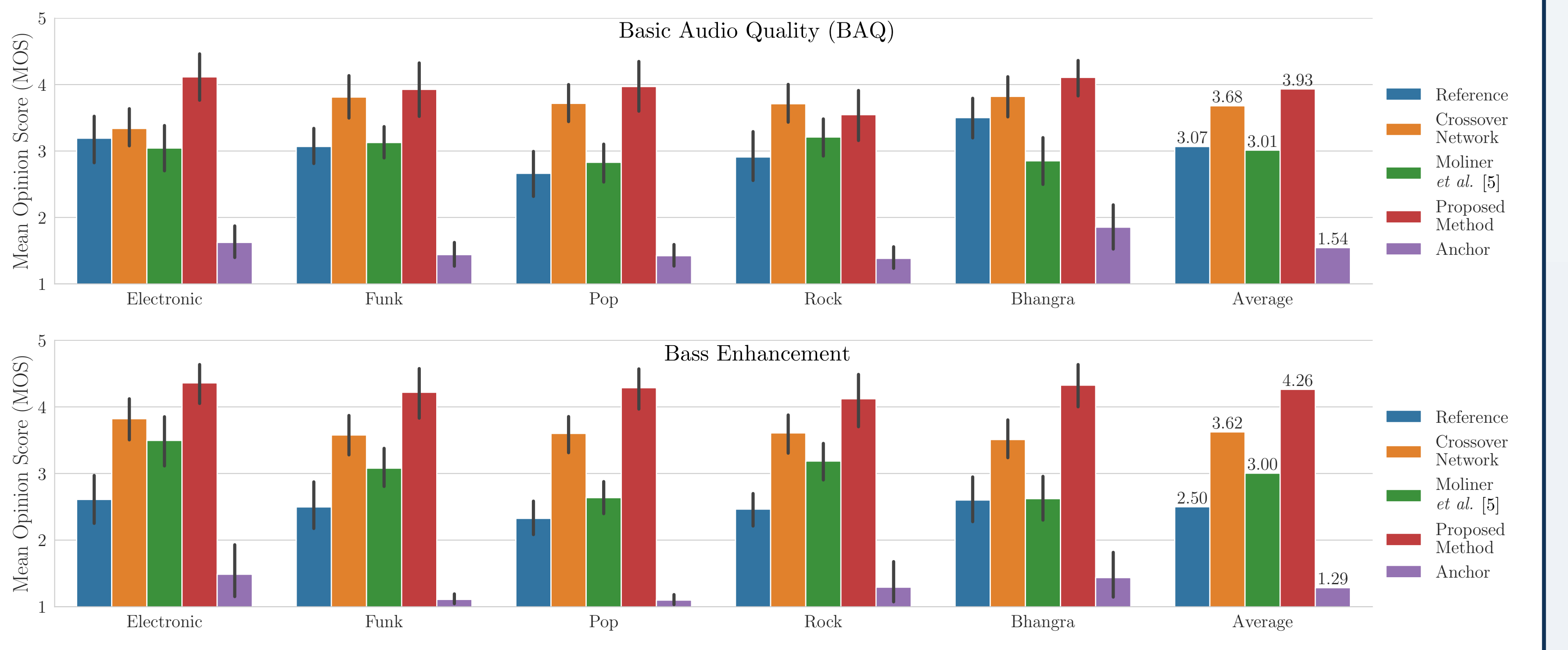
3. Perceptual Test

We run an experimental campaign consisting in two Mean Opinion Score (MOS) tests (ITU-T Rec. P.800.1.) The listeners were asked to rate excerpts of the songs listed in the table on a scale of 1 (Bad) to 5 (Excellent) as far as both Basic Audio Quality (BAQ) and bass enhancement are concerned. Altogether, 30 people participated to the perceptual test (with an average age of 26 years.)

All the tracks were highpassed at 200 Hz but the "Anchor" which is the highpassed version of the "Reference" at 500 Hz. In order to avoid biases due to the perceived loudness, we applied to all the items first a peak normalization at -1.0 dB and then a LUFS (Loudness Unit referenced to Full Scale) normalization at -14.0 dB LUFS (i.e., Spotify's setting recommendations.)

The songs included in the perceptual test are:

Song	Artist	Music Genre
Giorgio, By Moroder	Daft Punk	Electronic
Get up (I Feel Like Being A) Sex Machine	James Brown	Funk
Oops!... I Did It Again	Britney Spears	Pop
By The Way	Red Hot Chili Peppers	Rock
Mundian To Bach Ke	Punjabi MC	Bhangra



The proposed method is proved to be the best both as far as BAQ and bass enhancement are concerned.

Audio examples are available. Scan the QR code.

References

[1] R. Giampiccolo, A. I. Mezza, A. Bernardini, A. Sarti, "Virtual Bass Enhancement Via Music Demixing," *IEEE Signal Process. Lett.*, 2023.
[2] R. Hennequin, A. Khlif, F. Voituret, M. Moussallam, "Spleeter: A Fast and Efficient Music Source Separation Tool with Pre-Trained Models," *J. Open Source Softw.*, 2020.
[3] N. Oo, W. Gan, M. O. J. Hawksford, "Perceptually-Motivated Objective Grading of Nonlinear Processing in Virtual Bass Systems," *J. Audio Eng. Soc.*, 2011.
[4] R. Giampiccolo, A. Bernardini, A. Sarti, "A Time-Domain Virtual Bass Enhancement Circuital Model for Real-Time Music Applications," *IEEE MMSP*, 2022.
[5] E. Moliner, J. Rämö, V. Välimäki, "Virtual Bass System with Fuzzy Separation of Tones and Transients," *DAFx*, 2020.

**Riccardo Giampiccolo and Alessandro Ilic Mezza contributed equally to this work.*