Learned Continuous Synthesis of Quadratic Difference Tone Spectra

Esteban Gutiérrez
Department of Information and Communications Technologies
Universitat Pompeu Fabra
esteban.gutierrezc@upf.edu

Behzad Haki
Independent Researcher
Barcelona, Spain
behzad.haki@upf.edu

Christopher Haworth
Department of Music
University of Birmingham
c.p.haworth@bham.ac.uk

Xavier Serra
Department of Information and Communications Technologies
Universitat Pompeu Fabra
xavier.serra@upf.edu

Rodrigo F. Cádiz
Music Institute and Department of Electrical Engineering
Pontificia Universidad Católica de Chile
rcadiz@uc.cl

Abstract

Quadratic difference tones (QDTs) are a species of auditory distortion product in which a "phantom" pure tone, absent from the acoustic signal, is clearly audible to listeners. Exploiting this phenomenon, one can synthesize harmonically rich tones for musical purposes, a technique called Quadratic Difference Tone Spectrum (QDTS) synthesis. Previous works have introduced numerical methods to synthesize QDTS based on the distortion function, which links a target QDTS and an overtone-structured carrier signal. While accurate, these methods were stochastic and discontinuous, making them difficult to control for musical purposes and effectively limiting them to stationary signals. This paper proposes a neural network-based approach that learns an approximate inverse of the distortion mapping in an autoencoder-like configuration, producing a continuous approximation that addresses prior limitations. Experimental results show that, although slightly less numerically precise, the method is sufficient for perceptual and musical applications. We also implement a real-time version in Max and evaluate its performance. Various sound examples demonstrate its expressive and musical potential. The source code, audio examples, tutorials, and software accompanying this work are available at https://cordutie.github.io/projects/qdts.html.

1 INTRODUCTION

Quadratic difference tones (QDTs) are one of a family of so-called auditory distortion products (ADPs), sound sources that appear in the presence of particular acoustic stimuli but that are absent from the physical sound signal (i.e., they do not appear in spectrogram analyses). Historically known as difference and combination tones, ADPs are today best understood as subjective sounds generated by nonlinearities in the active components of the cochlea, specifically the outer hair cells and their interaction with the basilar membrane (Johnsen et al., 1983). They are typical of healthy hearing and their testing has become a common diagnostic tool for identifying hearing disorders (Kemp, 1978; Johnsen et al., 1983; Abdala and Visser-Dumont, 2001).

In hearing studies, ADPs are typically evoked by presenting two primary sinusoidal signals, f_1 and f_2, and asking subjects to report the perceived sonority. In these conditions, two ADPs are particularly salient: the QDT that is the subject of this study, and the lower cubic difference tone (LCDT). The LCDT occurs at $2f_1 - f_2$, obeys cubic nonlinearity, and its amplitude is highly sensitive to the frequency ratio of the primaries. It also exhibits the highest level when recorded as a DPOAE in the ear canal. The QDT occurs at $f_2 - f_1$, is governed by square-law distortion, and has a lower DPOAE level. It is less sensitive to ratio than the LCDT, and thus retains its amplitude across wider variations in interval size. Additionally, the QDT is easier to perceptually segregate than the LCDT (Dewey, 2022; Zwicker, 1979; Plomp, 1965).

QDTs have a rich musical history dating back to their discovery by violinist Giuseppe Tartini in 1754 (Nelson, 2003). Their use ranges from the transient "ghost" tones elicited by improvisers like John Butcher and Evan Parker to the experimental drone works of Maryanne Amacher, Phill Niblock, and Catherine Christer Hennix. A primary challenge of deploying QDTs for music synthesis has been the sound pressure levels required for them to be audible. Using a two-sinusoid stimulus requires the gain to be at levels that are uncomfortable for most listeners.

To ameliorate this problem, Haworth (2011) presented a method utilizing a complex of pure tones, building upon previously established psychoacoustic findings (Pressnitzer and Patterson, 2001). By arranging the frequencies so that each consecutive pair of acoustic sinusoids produces the identical QDT, this configuration boosts the overall distortion gain while simultaneously generating harmonic components of the primary QDT. Furthermore, increasing the number of acoustic tones and spreading them over a wider frequency range allows the subjective level of the acoustic stimulus to be reduced, which greatly diminishes listener fatigue. This technique, termed Quadratic Difference Tone Spectrum (QDTS) synthesis, is the focus of this article. Formalizing the approach, Kendall et al. (2014) proposed models for QDTS synthesis, which afford unprecedented control over the timbre of the resulting auditory illusion.

This shift from raw phenomenon to controllable synthesis model has spurred a recent wave of integration into contemporary electronic music and sound art. The precise architectural manipulation of QDTS has been featured in recent electroacoustic releases and installations, exemplifying the impact of the technique on current experimental practices. Notable applications include the microtonal inner ear orchestrations of Marcin Pietruszewski (Pietruszewski, 2024), procedural and iterative resynthesis deployments by Hecker and Pietruszewski (2025), and the dissociative synthetic vocal works of Francis and Tagen (2025).

Despite these creative successes, existing algorithmic implementations of QDTS synthesis suffer from severe limitations that have constrained their musical use. The original symbolic method of Kendall et al. (2014) was bottlenecked at four harmonics, beyond which closed-form solutions are not generally available. The numerical extension introduced by Gutiérrez et al. (2023; 2024) lifted this restriction by employing a stochastic Newton-Raphson solver capable of handling an arbitrary number of harmonics, but in doing so introduced a critical continuity problem. Because the underlying inverse problem is generally multi-valued, the solver selects a different local inverse at each time step with no mechanism to enforce consistency between successive solutions. When the target spectrum varies in time, as it must in any musically interesting context, the recovered carrier complex jumps between distant solution branches, producing audible discontinuities in the synthesized signal. In practice, this restricted prior tools to stationary or near-stationary targets, severely limiting their expressive range.

The present paper proposes a neural network-based solution to the continuity problem that renders dynamic QDTS synthesis viable as a robust, real-time tool for contemporary composers. The central idea is to replace stochastic root-finding with a learned, deterministic mapping from target spectra to carrier complexes. Concretely, we frame the inverse of the distortion function as the encoder of an autoencoder-like configuration in which the (analytically known) distortion function itself acts as a fixed decoder. Continuity is then a structural property of the learned map rather than a quantity that must be enforced post-hoc. To respect the degree-2 homogeneity of the distortion function, we factor the network architecture so that scaling behavior is exact by construction, leaving the network only to learn the angular component of the inverse on the unit sphere. Together, these design choices yield a solver that is continuous, deterministic, and inexpensive to evaluate.

The contributions of this paper are threefold. First, we introduce a continuous, neural approximation to the QDTS inverse problem, trained through a curriculum-style procedure that mitigates the multiplicity of local inverses by progressively expanding the sampling domain. Second, we provide a real-time Max external and a companion patch, making the method directly accessible to composers in a familiar environment. Third, we present a quantitative evaluation of the proposed solver against the prior Newton-Raphson baseline along three axes: reconstruction accuracy, control smoothness, and computational efficiency. Additionally, we complement these results with a set of sound examples demonstrating previously unattainable forms of expressive control, including continuous timbral morphing, resynthesis of time-varying monophonic audio, and explicit manipulation of carrier phase as a compositional parameter.

The remainder of the paper is structured as follows. Section 2 develops the mathematical formulation of the QDTS model and analyzes the perceptual constraints, feasibility issues, and continuity problem inherited from prior work. Section 3 introduces our neural solver, including its homogeneity-preserving architecture and curriculum-style training procedure. Section 4 describes the real-time Max implementation and presents the quantitative benchmarks. Section 5 discusses the accompanying sound examples, with attention to resynthesis and phase-dependent regimes. Section 6 concludes the paper and Section 7 outlines directions for future work, with particular emphasis on the open problem of extending the framework to cubic difference tones.

2 BACKGROUND

In this section we introduce the QDTS model using a formulation consistent with our proposed approach. We also discuss the intrinsic limitations and problems with the model, including perceptual constraints, mathematical feasibility, and controllability issues.

2.1 Quadratic Difference Tone Spectra in a Nutshell

The distribution of amplitudes of a Quadratic Difference Tone Spectrum (QDTS) has been theorized by Kendall et al. (2014) and further explored by Gutiérrez et al. (2023; 2024). For convenience in our formulation, we give a brief summary of the functional formulation proposed in the latter.

In the QDTS model, there are two signals: one acoustic signal representing the *carrier* complex of sinusoids, and one representing the perceptual auditory distortion product evoked by the carrier, acting as the *target* complex of harmonically distributed sinusoids. In this framework, the *carrier* complex is comprised of sinusoids of frequencies $C, C+F, C+2F, \ldots, C+NF$, where $C, F > 0$ and N is an integer; and the *target* QDTS generated by such complex is then comprised of N pure tones of frequencies $F, 2F, \ldots, NF$. Moreover, if the amplitudes of the carrier complex of sinusoids are $a_0, \ldots, a_N$, the amplitudes of the harmonics of the QDTS, respectively $t_1, \ldots, t_N$, are the result of the application of the *distortion function* $D : \mathbb{R}^{N+1} \to \mathbb{R}^N$ on the carrier distribution of amplitudes:

$$D(a_0, \ldots, a_N) = (t_1, \ldots, t_N), \tag{1}$$

where each target amplitude is computed by

$$\begin{aligned} t_1 &= a_0 a_N \\ t_2 &= a_0 a_{N-1} + a_1 a_N \\ &\vdots \qquad\qquad \vdots \\ t_{N-1} &= a_0 a_2 + a_1 a_3 + \cdots + a_{N-2} a_N \\ t_N &= a_0 a_1 + a_1 a_2 + \cdots + a_{N-2} a_{N-1} + a_{N-1} a_N \end{aligned} \tag{2}$$

See a graphic representation of this setup in Figure 1.

Hence in theory, by solving the equation $D(a) = t$ for a fixed vector t, one can determine the carrier needed to generate any given target harmonic signal. This is what gives this model its sound synthesis appeal.

This model is not flawless and is subject to the same perceptual constraints as the underlying QDTs. These sounds are known to vary in perceived strength depending on the frequency range and the ratios of the carrier signal. The audibility of the QDT depends strongly on the intensity of

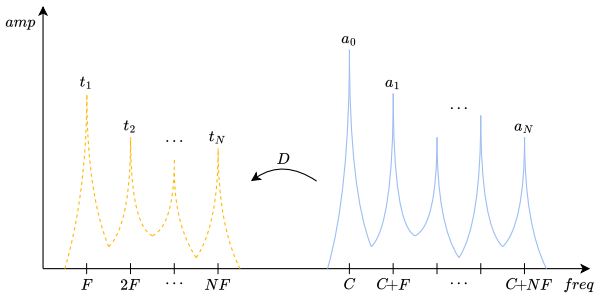

Figure 1: Illustrative frequency-amplitude representation of the QDTS model. On the right, the distribution of amplitudes of the carrier complex of pure tones is shown in cyan. On the left, the distribution of amplitudes of the target complex, evoked as a distortion product by the carrier, is shown dashed and in orange. The distortion function D is added to show how these signals are connected.

the acoustic tones: a level above 50 dB SPL is required for the component to be heard, and its amplitude then increases proportionally with the primaries, such that for every 1 dB increase, the QDT grows by 2 dB (Hartmann, 2004).

2.2 Feasibility and Workarounds

Beyond the perceptual constraints of QDTS, finding a definitive mathematical solution to Equation (2) presents significant challenges. Solving $D(a) = t$ requires finding the roots of a system of multivariate polynomials. While symbolic computation yields exact solutions for systems where $N \leq 4$ (Kendall et al., 2014), only the existence of complex solutions has been proven for orders $N = 5$ and $N = 6$ (Gutiérrez et al. 2023; 2024). Furthermore, simple examples lacking any real solution can easily be found for $N \geq 5$, making the problem analytically unsolvable in the general case.

Previous research (Gutiérrez et al. 2023; 2024) bypassed this limitation by employing a stochastic numerical approach. By applying the Newton–Raphson algorithm, an approximate solution could be calculated. If the algorithm failed to converge within a set number of iterations, a microscopic random perturbation was added to the target distribution, and the process was restarted. Even though there was no theoretical guarantee for an algorithm of this nature to converge, empirical testing demonstrated that this approach yielded an approximate solution that is still perceptually relevant in over 99% of cases. Ultimately, this approach proved that despite the lack of formal mathematical certainty, computing approximate solutions remains a highly feasible workaround for the problem.

2.3 The Continuity Problem

The distortion function D, like any multi-variable C^∞ function, has the potential to admit many local inverses in different parts of its co-domain. In practice, this implies that given a target spectrum, the system of equations (2) generally may admit multiple real solutions, and therefore an infinite set of approximate solutions.

This multiplicity becomes problematic when the target spectrum evolves over time. Numerical and stochastic solvers, such as the method introduced by Gutiérrez et al. (2023; 2024), effectively select one of these local inverses at each step, but provide no mechanism to enforce consistency between successive solutions. As a result, the recovered carrier complex may jump between distant solution branches, producing abrupt changes in the carrier signal. These discontinuities manifest as audible glitches or artifacts, and up to this contribution, remained as an open problem.

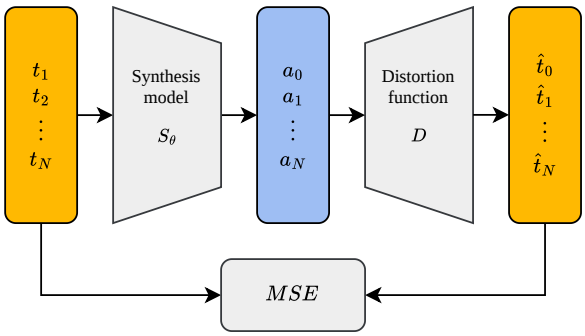

Figure 2: Autoencoder-like architecture. The distortion function works as a decoder, while the synthesis function is learned.

3 LEARNED INVERSE DISTORTION FUNCTION

In essence, solving the QDTS model amounts to finding a map that transforms target spectra into carrier complexes. We call this map the *synthesis function* $S : \mathbb{R}^N \to \mathbb{R}^{N+1}$. Ideally, S should act as a left inverse of the distortion function, meaning that it produces outputs whose distortion matches the target spectrum:

$$D(S(t)) = t. \tag{3}$$

Moreover, in order to address continuity issues, S must also be continuous.

As discussed in Subsection 2.2, the relation (3) may not always be achievable; however, the relaxation $D(S(t)) \approx t$ is feasible and still perceptually relevant.

In this section, we propose to learn a map $S = S_\theta$ with these properties using a shallow neural network, and to enforce such properties through simple architectural and training choices.

3.1 Architecture

The distortion function is homogeneous of degree 2, i.e., for $\lambda \in \mathbb{R}$ and $a \in \mathbb{R}^{N+1}$,

$$D(\lambda a) = \lambda^2 D(a). \tag{4}$$

It follows that if S is a left inverse of D, it must satisfy the corresponding inverse relation

$$S(\lambda t) = \sqrt{\lambda}\, S(t), \tag{5}$$

for $t \in \mathbb{R}^N$ and $\lambda > 0$.

We enforce this property by construction:

$$S_\theta(t) = \sqrt{\|t\|}\, s_\theta\left(\frac{t}{\|t\|}\right), \tag{6}$$

where $t \neq 0$ and s_θ is a neural network.

This construction not only ensures Equation (5), but it also guarantees continuity, improves training stability by operating exclusively on normalized vectors, and allows

the function to extend naturally to the full domain once trained on the N-dimensional sphere.

Given the low complexity of the problem and the lack of an obvious relationship between input dimensions, a natural baseline is a shallow Multi-Layer Perceptron (MLP) directly modeling s_θ. Adopting this baseline, empirical testing revealed that a three-layer MLP yielded the best results.

3.2 Training

This formulation can be interpreted as an autoencoder-like architecture in which the distortion function D acts as a fixed decoder mapping acoustic signals to psychoacoustic phenomena, while S_θ serves as a learned encoder performing the inverse mapping. The model is trained by minimizing the reconstruction loss

$$\mathcal{L}(t) = \|D(S_\theta(t)) - t\|^2. \tag{7}$$

See an illustration of this autoencoder-like approach in Figure 2.

As noted earlier, some regions of the codomain of D may admit multiple inverses, which can complicate training. To mitigate this, we train on random batches sampled from spheres of increasing radius. This encourages the model to first learn a local inverse and then extend it smoothly to larger domains.

Specifically, we begin with $10{,}000$ batches of 512 random target vectors sampled from a sphere centered at $(0.5, \ldots, 0.5)$ with radius 0.1, and progressively increase the radius in steps of 0.1 up to 1.

Due to non-convexity and the presence of multiple possible inverses, we perform 8 runs with different initializations for each $N \in \{5, \ldots, 16\}$ and another 8 runs using a regularizer that makes $a_0 \approx t_1$, a constraint also used by Gutiérrez et al. (2023; 2024) that helps localize the solutions. Although this procedure provides no formal guarantees, it consistently yields accurate and continuous approximate inverses as shown in Section 4.2.

4 IMPLEMENTATION IN MAX AND BENCHMARKS

In this section, we introduce a real-time implementation of our neural network-based QDTS synthesizer, alongside a companion patch designed to seamlessly integrate this technique into compositional workflows. To better understand the performance and quality of the developed external, we also conduct a series of quantitative experiments. The objective of these experiments is to evaluate the quality of the reconstructed target spectra, the smoothness of the output under continuous parameter changes, and the computational efficiency of our implementation in a real-time setting.

4.1 Max Implementation

To allow composers to interact with the trained models in a familiar environment, we developed a custom Max external called `qdts.solver_nn`, as well as a companion Max patch (Cycling '74, 2024; Puckette and Zicarelli, 1990).[1]

The external uses ONNX Runtime (2021) to run inference on the pre-trained models, exported in the ONNX format (Bai et al., 2019). This approach results in a self-contained package that is straightforward to install in a manner with which Max users are familiar.

The patch (see Figure 3) allows the user to specify the target and carrier frequencies and the number of harmonics, and to interactively adjust the target amplitude distribution in real time. Based on the configuration, `qdts.solver_nn` automatically selects a compatible model and converts the target amplitudes to the corresponding carrier amplitudes, which are then passed to a custom additive synthesizer to auralize the resulting spectrum. Moreover, the user can choose any of the 16 models trained for each case, giving further possibilities to the timbre of the acoustic signal while generating the same QDTS.

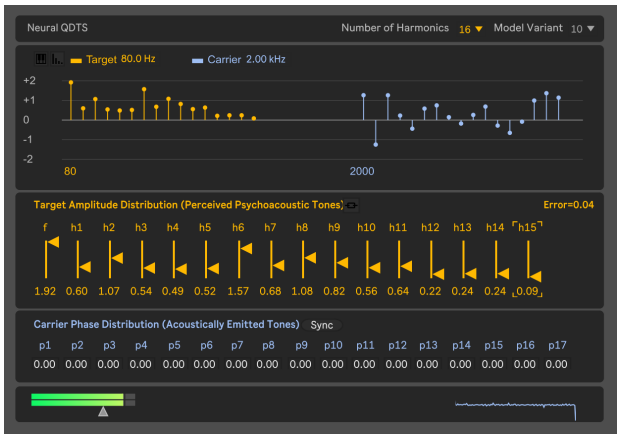

Figure 3: A snapshot of the developed Max patch to allow for easy interaction with the `qdts.solver_nn` external employing the trained models.

4.2 Generation Quality

To assess the quality of the generations, we evaluated the reconstruction error of the deployed models across all supported target spectrum sizes $N \in \{5, \ldots, 16\}$ using 10,000 random target amplitude vectors drawn uniformly from $[0, 1]^N$. For each input, reconstruction error is measured by the same loss function used during training, $\mathcal{L}$ in Equation (7), to quantify how closely the distortion products of the predicted carrier match the target spectrum.

For reference, we also compare against the Newton–Raphson implementation of Gutiérrez et al. (2023; 2024), evaluating the same metric. Figure 4 shows the mean and standard deviation of $\mathcal{L}$ as a function of N for both solvers.

As shown in Figure 4, the Newton–Raphson solver achieves lower reconstruction error across all values of N. Nevertheless, the neural network solver remains well within an acceptable range, with mean errors below 0.04 across all supported spectrum sizes, confirming that it reliably approximates the inverse of the distortion function throughout the supported range. The narrow variance band indicates that performance is consistent across model variants and random inputs, suggesting that the deployed models yield stable solutions.

4.3 Control Smoothness

To evaluate whether `qdts.solver_nn` produces smooth outputs under continuous parameter changes, we measured how closely the model's response to a linearly interpolated

[1] The external, patch, and the source codes are openly accessible from https://cordutie.github.io/projects/qdts.html.

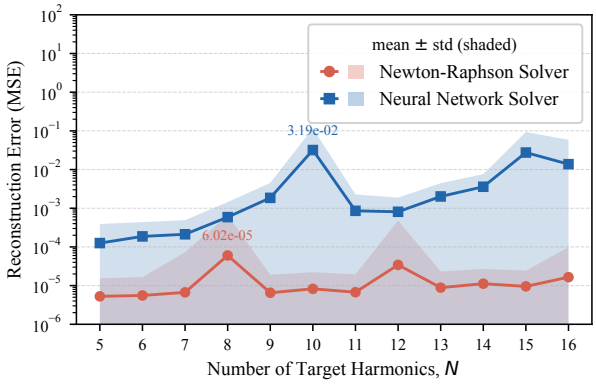

Figure 4: Reconstruction error as a function of N for the neural network solver and the Newton–Raphson solver. For each N, the results are aggregated across all 16 model variants.

input follows a linear interpolation of the endpoint outputs. For each target spectrum size $N \in \{5, \ldots, 16\}$, we generated 10,000 random endpoint pairs $\mathbf{A} \sim \mathcal{U}[0,1]^N$ and $\mathbf{B} = 1 - \mathbf{A}$, and queried the model at 11 evenly spaced interpolation points $\mathbf{T}_\alpha = (1-\alpha)\mathbf{A} + \alpha\mathbf{B}$, $\alpha \in \{0, 0.1, \ldots, 1.0\}$, across all 16 model variants. At each α, smoothness is quantified by the normalized output distance

$$r(\alpha) = \frac{\|S_\theta(\mathbf{T}_\alpha) - S_\theta(\mathbf{A})\|}{\|S_\theta(\mathbf{B}) - S_\theta(\mathbf{A})\|}, \tag{8}$$

which equals α for a perfectly linear model.

Figure 5 shows the distribution of $r(\alpha)$ pooled across all N and model variants. This analysis is specific to the neural network solver; an equivalent evaluation of the Newton–Raphson solver is not meaningful, as it selects solutions independently at each time step with no mechanism to enforce consistency between successive outputs, and would therefore produce arbitrary $r(\alpha)$ values regardless of input smoothness.

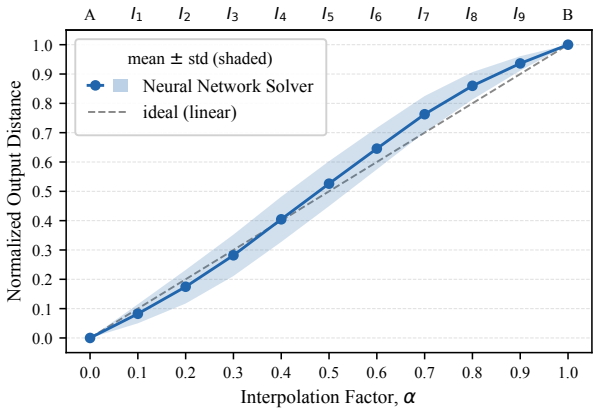

Figure 5: Normalized output distance $r(\alpha)$ of the neural network solver at 11 interpolation points between two random endpoint spectra, pooled across all target spectrum sizes $N \in \{5, \ldots, 16\}$ and all 16 model variants. The dashed diagonal represents the ideal linear response $r(\alpha) = \alpha$.

As shown in Figure 5, `qdts.solver_nn` produces smooth, nearly linear outputs under continuous input interpolation. The normalized output distance $r(\alpha)$ tracks the ideal diagonal closely across all interpolation points, indicating that the carrier tone amplitudes transition continuously and predictably as the target spectrum is varied, with only a small degree of nonlinearity. The consistently narrow variance band across all values of N and model variants confirms that this smooth behavior is stable and not specific to particular inputs or configurations.

4.4 Computational Efficiency

The computational efficiency of `qdts.solver_nn` was evaluated by measuring the execution time of the Max external[2] across all supported target spectrum sizes $N \in \{5, \ldots, 16\}$ using 10,000 random amplitude sequences. For reference, the same benchmark was applied to the Newton–Raphson implementation of Gutiérrez et al. (2023; 2024). Figure 6 summarizes the results.

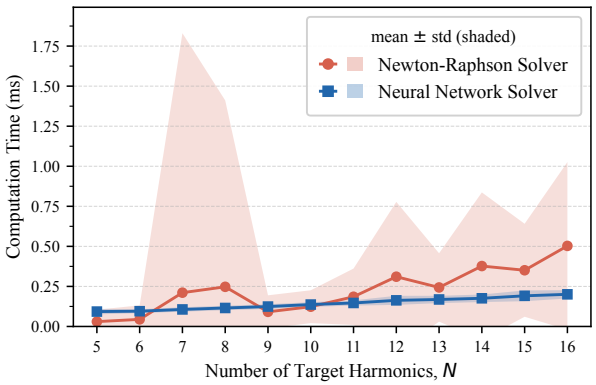

Figure 6: Computation time of Newton–Raphson solver and neural network solver. Neural network solver results are aggregated across all 16 model variants.

As shown in Figure 6, `qdts.solver_nn` achieves a smooth, predictable increase in computation time with N, whereas the Newton–Raphson implementation exhibits higher and more irregular times across all sizes. The stochastic nature of the Newton–Raphson restart mechanism results in high variance, particularly for ill-conditioned inputs that require many restarts before convergence. In contrast, `qdts.solver_nn` executes a fixed sequence of neural network operations regardless of the input, yielding near-constant variance across all values of N. Notably, the computation time of `qdts.solver_nn` remains below 0.25 ms for all values of N, demonstrating its suitability for real-time audio applications.

5 SOUND EXAMPLES

This section presents a series of sound examples demonstrating the neural network-based implementation of QDTS synthesis. The examples are designed to illustrate both the robustness of the learned model across different pitch conditions and its behavior in resynthesis, interaction with acoustic signals, and phase-dependent regimes. All audio examples are available in the supplementary material repository[3]. In all cases, synthesis is performed using our

[2] Measured on an Apple M4 Pro.

[3] https://cordutie.github.io/projects/qdts.html

proposed neural network solver. To achieve the spatial and perceptual effects typical of QDTs, listeners are advised to use loudspeakers at moderate to high playback levels rather than headphones.

5.1 Basic Synthesis Examples

In this first set of examples, we demonstrate stationary pitch synthesis using our method. The objective is to evaluate the ability of the neural solver to generate stable QDTS structures while allowing continuous control over its harmonic content, a previously impossible feature.

Each example consists of a fixed-pitch synthetic tone generated via additive synthesis, which is then used as a target representation for the neural solver. The model outputs a corresponding carrier complex, with a different number of harmonics in each case.

We repeat this procedure for three distinct fundamental frequencies using three independently trained models, demonstrating that the learned mapping generalizes across pitch space and does not depend on a single configuration. The configurations used in each example are detailed below, where $\mathcal{M}_m^n$ denotes the m-th model variant trained for the $N = n$ harmonic case:

- Example A: 9 harmonics with fundamental pitch $F = 73.4$ Hz, using carrier $C = 2$ kHz, and neural network model $\mathcal{M}_2^9$.
- Example B: 10 harmonics with fundamental pitch $F = 98.0$ Hz, using carrier $C = 2.5$ kHz, and neural network model $\mathcal{M}_7^{10}$.
- Example C: 14 harmonics with varying fundamental pitch, using carrier $C = 2.5$ kHz, and neural network model $\mathcal{M}_{11}^{14}$.

Across all cases, our model output preserves stable perceptual pitch while exhibiting controlled redistribution of harmonic energy through the learned synthesis function.

5.2 Resynthesis Examples

This subsection presents QDTS resynthesis of real-world monophonic audio recordings using joint pitch tracking and harmonic distribution estimation. Although previous models have been applied to similar tasks, the lack of continuity control forced them to rely on constant harmonic distributions, a strong constraint that is lifted by our continuous model.

We first exemplify this approach using the same recording used by Gutiérrez et al. (2023) for the purpose of comparison and add a second sound example. Here, a pitch estimator and harmonic tracking module estimate time-varying fundamental frequencies and spectral envelopes, which are then fed into the neural solver to reconstruct a corresponding QDTS representation. While demonstrated here in an offline setting, the pipeline is equally applicable in a real-time context, subject to the latency constraints of frame-based analysis.

5.3 Interactions Between Distortion Products and Acoustic Signals

This set of examples explores perceptual interactions between QDTS-generated distortion products and externally introduced acoustic sinusoids. The goal is to examine beating effects and interference phenomena arising when external tones are positioned near predicted distortion frequencies.

A long-form interactive example demonstrates this by introducing an additional acoustic sinusoid near the frequency of each of the first three harmonics. The listener is encouraged to notice how, despite the absence of acoustic interactions between the physical signals, audible beating occurs due to interference with the perceptual distortion products generated by our model.

5.4 Phase Interactions

The final set of examples investigates the role of carrier phase configuration in QDTS synthesis. While harmonic magnitudes remain fixed, variations in phase relationships of the carrier complex significantly affect the resulting distortion perceptibility.

A stationary pitch is used throughout, while the phase vector of the carrier is systematically modified. In particular, configurations are selected that either enhance or suppress the perceptual prominence of the distortion products.

- Phase configuration A: Enhanced QDTS emergence.
- Phase configuration B: Suppressed QDTS emergence.
- Phase sweep experiment (continuous morphing).

The perceptual prominence of the distortion products is highly dependent on the phase configuration of the carrier signal. Users should therefore be aware that identical magnitude targets may yield markedly different perceptual outcomes depending on the carrier phase. For this reason, the Max patch exposes direct control over individual phase parameters, allowing the performer to explore and exploit these interactions in real time.

6 CONCLUSIONS

This paper has presented a neural network-based approach to the synthesis of Quadratic Difference Tone Spectra that resolves the continuity limitations of prior numerical solvers and renders dynamic QDTS synthesis viable for real-time musical use. By framing the inverse of the distortion function as the encoder of an autoencoder-like configuration, with the analytically known distortion function acting as a fixed decoder, and by enforcing the degree-2 homogeneity of the underlying mapping at the architectural level, we obtained a continuous, deterministic approximate inverse that can be evaluated cheaply at audio-rate control timescales.

Our quantitative benchmarks indicate that the neural solver trades a modest amount of numerical accuracy for a set of properties that prior solvers could not provide simultaneously. Compared with the stochastic Newton–Raphson baseline of Gutiérrez et al. (2023; 2024), the proposed model exhibits slightly higher reconstruction error, yet mean errors remain below 0.04 across all supported harmonics, an acceptable range for perceptual listening. In exchange, the neural solver produces near-linear responses to continuous parameter interpolation, and it executes comfortably within the budget of real-time audio applications. Crucially, both properties are structural rather than incidental: continuity follows from the architecture, and computational cost is independent of the input.

These properties are reflected in the accompanying sound examples, which include stable stationary synthesis across

multiple fundamentals and trained models, resynthesis of time-varying monophonic recordings with continuously evolving harmonic envelopes, controlled interactions between predicted distortion frequencies and externally introduced acoustic sinusoids, and phase-dependent regimes in which identical magnitude targets yield markedly different perceptual outcomes. Several of these scenarios, in particular continuous timbral morphing and time-varying resynthesis with unconstrained harmonic distributions, were not practically achievable with prior solvers. The Max external `qdts.solver_nn` and its companion patch make these capabilities directly available to composers in a familiar environment.

7 FUTURE WORK

Several directions for future work follow naturally from the framework presented here. The most demanding concerns Cubic Difference Tones (CDTs), and the question of whether a coherent "Cubic Difference Tone Spectra" (CDTS) framework can be formulated at all. CDTs, and in particular the lower CDT at $2f_1 - f_2$, are perceptually richer than QDTs in several respects: they are typically more salient at moderate stimulus levels, they are the dominant distortion product measured in clinical distortion product otoacoustic emission (DPOAE) protocols, and they exhibit a more pronounced phenomenological separation from the acoustic primaries. From a compositional standpoint, this makes them an attractive target; from a modeling standpoint, however, the problem differs from the quadratic case in ways that prevent a direct transposition of the methodology presented here.

Three obstacles stand out. First, the underlying nonlinearity is cubic rather than quadratic, so the analogue of the distortion function D is homogeneous of degree three. While this is relatively simple to enforce architecturally, it raises the polynomial degree of the system relating carrier and target amplitudes, increasing the multiplicity of local inverses. Second, and more fundamentally, the perceptual amplitude of the LCDT depends strongly and non-monotonically on the frequency ratio f_2/f_1 of the primaries (Plomp, 1965; Zwicker, 1979). The QDTS framework exploits the fact that the quadratic component is approximately invariant to interval size and adds linearly across components of a harmonic complex with constant difference frequencies; no comparable invariance holds for the cubic component, and any CDTS distortion function would have to incorporate a frequency-ratio-dependent gain term whose form is not yet well established. Third, the LCDT is more sensitive than the QDT to absolute level, primary level imbalance, and individual cochlear variability, weakening the population-level validity of any synthesis target.

Two lower-risk directions concern the QDTS framework itself. The current solver inverts only the magnitude relationship between carrier and target, but as noted in Section 5, the perceptual prominence of QDTs depends substantially on carrier phase. Incorporating phase as a controllable target dimension, either by training joint magnitude–phase inverses or by learning auxiliary maps that select phase configurations optimizing perceptual prominence, would tighten the link between the model's mathematical targets and the listener's experience. Separately, the training procedure used here is deliberately simple, and more sophisticated schemes, including approaches that explicitly account for the multiplicity of local inverses by guiding the network toward a designated branch, could further improve reconstruction accuracy without compromising the structural properties on which the present method relies.

ACKNOWLEDGEMENTS

This research was supported by the ANID Fondecyt Regular Grant #1230926, funded by ANID Millenium Nucleus NCS2025-12 ANIMUPA, Government of Chile, and the project "Cátedra en IA y Música" (TSI-100929-2023-1), funded by the Secretaría de Estado de Digitalización e Inteligencia Artificial, the European Union-Next Generation EU funds and BMAT Music Innovators.

We are profoundly grateful to Marcin Pietruszewski for his support and interest in our research. His pioneering work has continually inspired this project. Furthermore, his artistic practice and research, including his SuperCollider ports[4], have served as a pillar in the dissemination of these niche but fascinating synthesis techniques to a broader audience.

REFERENCES

Abdala, C. and Visser-Dumont, L. (2001). Distortion product otoacoustic emissions: A tool for hearing assessment and scientific study. *The Volta review*, 103(4):281–302.

Bai, J., Lu, F., Zhang, K., et al. (2019). Onnx: Open neural network exchange. https://github.com/onnx/onnx.

Cycling '74 (2024). Max 9. https://cycling74.com/products/max. Version 9.

Dewey, J. B. (2022). Cubic and quadratic distortion products in vibrations of the mouse cochlear apex. *JASA Express Letters*, 2(11):114402.

Francis, R. and Tagen, J. A. (2025). H e l l o after-person. Audio release. https://www.etat.xyz/release/AfterPerson.

Gutiérrez, E., Haworth, C., and Cádiz, R. (2023). Generating quadratic difference tone spectra for auditory distortion synthesis. In *Proceedings of the International Computer Music Conference*, pages 237–243.

Gutiérrez, E., Haworth, C., and Cádiz, R. F. (2024). Generating sonic phantoms with quadratic difference tone spectrum synthesis. *Computer Music Journal*, 47(3):19–34.

Hartmann, W. M. (2004). *Signals, sound, and sensation*. Springer Science & Business Media, New York.

Haworth, C. (2011). Composing with absent sound. In *Proceedings of the International Computer Music Conference 2011*, pages 342–345, University of Huddersfield, UK.

Hecker, F. and Pietruszewski, M. (2025). Normification. Audio release. https://normification.bandcamp.com/album/normification-diag071.

Johnsen, N. J., Bagi, P., and Elberling, C. (1983). Evoked acoustic emissions from the human ear. iii. findings in neonates. *Scandinavian Audiology*, 12(1):17–24.

[4] https://github.com/mpietrus00/sc-qdts

Kemp, D. T. (1978). Stimulated acoustic emissions from within the human auditory system. *The Journal of the Acoustical Society of America*, 64(5):1386–1391.

Kendall, G., Haworth, C., and Cádiz, R. F. (2014). Sound synthesis with auditory distortion products. *Computer Music Journal*, 38(4):5–23.

Nelson, S. M. (2003). *The Violin and Viola: History, Structure, Techniques*. Courier Dover Publications.

ONNX Runtime (2021). https://onnxruntime.ai/. Version: 1.20.0.

Pietruszewski, M. (2024). Distortion product lattice (fokker qdts wavepackets). Sound installation, Sonic Acts. https://sonicacts.com/archive/distortion-product-lattice.

Plomp, R. (1965). Detectability threshold for combination tones. *The Journal of the Acoustical Society of America*, 37(6):1110–1123.

Pressnitzer, D. and Patterson, R. (2001). Distortion products and the perceived pitch of harmonic complex tones. In Breebart, D., Houtsma, A., Kohlrausch, A., Prijs, V., and Schoonoven, R., editors, *Physiological and Psychophysical Bases of Auditory Function*, pages 97–104. Shaker Publishing BV, Maastricht, The Netherlands.

Puckette, M. and Zicarelli, D. (1990). Max-an interactive graphic programming environment. *Opcode Systems, Menlo Park, CA*.

Zwicker, E. (1979). Different behaviour of quadratic and cubic difference tones. *Hearing Research*, 1(4):283–292.