The IEEE-IS^2 2025 Music Packet Loss Concealment Challenge

Alessandro Ilic Mezza
DEIB
Politecnico di Milano
Milan, Italy
alessandroilic.mezza@polimi.it

Alberto Bernardini
DEIB
Politecnico di Milano
Milan, Italy
alberto.bernardini@polimi.it

Claudia Rinaldi
CNIT
Research Unit of L'Aquila
L'Aquila, Italy
claudia.rinaldi@univaq.it

***Abstract*—We present the IEEE-IS^2 2025 Music Packet Loss Concealment Challenge. Building on the foundations laid in the inaugural edition, this second installment of the challenge introduces improved evaluation metrics and a newly curated test set. In this paper, we outline the challenge rules, detail the construction of the validation and blind test sets, discuss the evaluation methodology used to benchmark participant submissions, and present the results. Continuing to bridge the communities of signal processing, machine learning, and networked music performance, the IEEE-IS^2 2025 Music Packet Loss Concealment Challenge further advances the state of the art in packet loss concealment for music signals and promotes the development of robust, perceptually-aware algorithms suitable for real-world immersive audio applications.**

***Index Terms*—audio packet loss concealment, networked music performance, networked immersive audio**

I. Introduction

The Internet of Sounds (IoS) [1] envisions seamless, high-fidelity, real-time communication of musical and audio experiences across distributed networks, blending sound and music computing with Internet of Things (IoT) paradigms. Central to this vision is Networked Music Performance (NMP), where geographically distributed musicians rehearse, improvise, and perform together over packet-switched networks as if they were sharing the same physical space [2]. Achieving this illusion requires ultra-low end-to-end latency, typically below 20–30 ms [3], to preserve the tight temporal coordination essential for ensemble playing. While even small timing deviations can impairs performers' ability to synchronize, network jitter can further delay a packet past its scheduled playback time [2].

Meeting these stringent latency requirements often means relying on best-effort transport protocols such as UDP, which, unlike connection-oriented protocols, forego built-in reliability mechanisms of like delivery guarantees and congestion control in order to minimize delay [2], [4]. While this choice helps maintain latency low, it inherently increases the likelihood of packet loss. As a result, lost packets remain unrecoverable (e.g., by retransmission), intensifying the challenge of preserving audio quality and seamless interaction in NMP systems. Packet loss indeed produce audible artifacts that risk breaking the continuity of musical phrases and reduce the sense of co-presence between remote participants.

Although Packet Loss Concealment (PLC), i.e., the ensemble of algorithms aiming to mitigate these degradations by data insertion of by synthesizing plausible reconstructions of missing audio segments, has been widely studied in the context of speech [8]–[10], music introduces greater technical and perceptual complexity. Musical audio often contains long-sustained tones, rich harmonic structures, rapid transients, and expressive variations in dynamics and articulation. These characteristics make networked music more sensitive to concealment errors compared to speech communications, especially under bursty loss conditions. The stringent latency constraints of NMP preclude the use of non-causal or high-look-ahead algorithms, meaning that PLC solutions must operate in real time with only past information available [4], [11], [12]. This combination of strict delay budgets, high perceptual quality requirements, and diverse signal characteristics makes music-oriented PLC a uniquely challenging problem.

Responding to this need, the 2nd IEEE-IS^2 Music Packet Loss Concealment Challenge[1] was organized as a satellite event of the 6th IEEE International Symposium on the Internet of Sounds (IEEE IS^2 2025), continuing the inaugural 2024 edition [13]. It aims to promote advances in PLC methods—

TABLE I
Team Ranking

	Average score $\pm$ SD	Median score	Trials won	Ranking
Aironi et al. [5]	68.82 ± 19.18	73.0	8	1st
Daniotti et al. [6]	57.66 ± 20.30	56.5	1	2nd
Severi [7]	52.01 ± 18.24	50.0	1	3rd
PARCnet-IS^2 (Baseline)	53.96 ± 20.27	55.5	0	–
Zero-filling (Anchor)	15.75 ± 14.49	11.5	0	–

This work was partially supported by the European Union—Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP D43C22003080001, partnership on "Telecommunications of the Future" (PE00000001—program "RESTART").

The organizers would like to thank Sennheiser Electronic SE & Co. KG for the generous sponsorship.

[1]**Official website:** https://internetofsounds2025.ieee-is2.org/workshops/3rd-ieee-international-workshop-networked-immersive-audio/music-packet-loss-concealment; **GitHub repository:** https://github.com/polimi-ispl/2025-music-plc-challenge.

supporting immersive musical performances in real-world networked scenarios.

Key enhancements in the 2025 edition include: (i) a richly diversified blind test set spanning acoustic instruments, electric instruments, synthesizers, and vocals; (ii) stricter enforcement of real-time, causal operations, pushing models to handle the operating conditions typical of networked immersive audio; (iii) a refined evaluation protocol, combining new objective metrics with well-established listening tests.

This paper presents an overview of the challenge, including its rules, datasets, baseline system, evaluation methodology, and the performance of participating teams. In doing so, it seeks to provide both a snapshot of the current state of the art and a roadmap for future research in music-oriented PLC.

The reminder of this manuscript is organized as follows. In Section II we review the challenge rules. In Section III, we outline the team submissions. In Section IV,we describe the evaluation protocol and metrics. In Section V, we present the challenge results. Finally, Section VI concludes the paper.

II. Challenge Overview

A. Challenge Rules

The IEEE-IS^2 2025 Music Packet Loss Concealment Challenge officially began on February 10, 2024. The blind test set was made available on June 12, 2025, and the submission window closed on June 20, 2025. Unlike the 2024 edition, where each team could submit up to two systems, the 2025 edition limited each team to a single submission.

Participants were required to design PLC systems operating at a sampling rate of 44.1 kHz, capable of concealing packet losses of 512 samples (11.6 ms). Although shorter packet sizes are often favored in NMP applications, the decision to use 512-sample packets was intended to present a more challenging test scenario, encouraging participants to address more complex conditions involving longer gaps.

To reflect the stringent latency demands of real-time musical collaboration over networks, all PLC systems were mandated to operate under real-time constraints. Furthermore, only causal systems were accepted, meaning predictions could only rely on past data, with no access to future (or lost) information. Unlike other audio PLC challenges [9], [10], the IEEE-IS^2 2025 Music Packet Loss Concealment Challenge prohibited the use of look-ahead. Nevertheless, techniques that manipulate future packets (e.g., cross-fading between reconstructed and valid segments) were allowed, provided that such information was not exploited to inform the prediction of missing audio data.

Beyond these constraints, no restrictions were placed on the types of PLC methods. Participants were free to leverage traditional signal processing techniques, deep learning models, or any hybrid combination thereof.

As in the previous edition, no training data was provided by the organizers, nor was a list of approved datasets specified. Nonetheless, participants were required to use only publicly-available and freely-accessible datasets for training purposes. Data augmentation was permitted without restriction, provided that models remained blind to any metadata or auxiliary information beyond packet loss traces.

To support development, we released a *validation set* (with ground truth), while for the final evaluation, we provided a new *blind test set* (without ground truth).

B. Validation Set

As validation data, we released the ground truth of the test set from the IEEE-IS^2 2024 Music Packet Loss Concealment Challenge.[2]

Thus, the 2025 validation set comprises 162 monophonic audio files in a 16bit-44.1kHz WAV format. These recordings were sourced from the publicly available AVAD-VR dataset [14], which contains anechoic audio and 3D-video captures of small ensemble performances. The dataset primarily includes classical and jazz pieces, showcasing a range of acoustic instruments such as violin, cello, clarinet, saxophone, double bass, and classical guitar.

Each audio file was segmented into 11.6-second clips using non-overlapping windows. Segments containing more than 30% silence were discarded to exclude audio signals with negligible energy over time. The remaining were normalized according to Rec. ITU-R BS.1770-4 using `pyloudnorm` [15].

To simulate packet loss, the audio clips were artificially corrupted using predefined packet traces, i.e., text files containing binary sequences where `0` indicates a correctly received packet and `1` denotes a lost packet. Each binary digit corresponds to 512 samples, and the traces imply packet timing based on the sampling rate; no explicit timing data is embedded in the traces.

The packet traces used for generating the blind test set were adapted from those released for the INTERSPEECH 2022 Audio Deep Packet Loss Concealment Challenge [9].[3] These traces are real-world measurements from actual network conditions. They are categorized into three subsets based on the maximum length of burst losses:

- **Subset 1:** Bursts of up to 6 consecutive packets;
- **Subset 2:** Bursts ranging from 6 to 16 packets;
- **Subset 3:** Bursts between 16 and 50 packets.

Intentionally excluding Subset 3 to avoid extreme conditions, we sampled traces from Subset 1 (with 90% probability) and Subset 2 (with 10% probability). Whenever necessary, multiple packet traces were concatenated in order to match the duration of the clean audio clip.

C. Blind Test Set

We base the design of the blind test set on that of the validation dataset. Namely, We simulate packet losses using the same very procedure (see Section II-B).

The main difference lies in the increased number of sound sources, which now include not only acoustic instruments but also electric instruments and synthesizers. As a result, the new blind test dataset contains 674 samples—more than four times the size of the 2024 corpus.

[2]Available: https://github.com/polimi-ispl/2025-music-plc-challenge

[3]Available: https://github.com/microsoft/PLC-Challenge

Single-instrument audio tracks are sourced from a range of publicly available datasets. While the names of these datasets are not disclosed, below we report the salient details about data selection and preprocessing, which differ from instrument to instrument.

- **Piano:** We select 59 stereophonic solo piano recordings of classical music pieces, each with a duration of 12 s.
- **Orchestral samples:** Anechoic solo recordings of bassoon, horn, viola, violin, flute, double bass, clarinet, cello, oboe, and trumpet are sourced from an open dataset of multi-track orchestral music. Each clip is manually segmented in a DAW (Reaper by Cockos Inc.) so as to encompass a coherent music phrase.
- **Synth:** We downloaded publicly-available synthesizer tracks from `freesound.org`, all licensed under Creative Commons. The search results were filtered and using the following parameters: Category: "Music", Subcategory: "Solo instrument", Type: "WAV", Sample rate: "44100", Bit depth: "16", and Tag: "loop". We select 69 of the first audio samples sorted using "Rating (highest first)" as a criterion. The resulting clips include both mono and stereo formats. Clips longer than 20 seconds are trimmed, while those shorter than 10 seconds are repeated twice.
- **Electric bass:** A total of 17 monophonic recordings of electric bass showcasing different playing techniques (fingerstyle, pick, slap) were first peak normalized (-1.0 dB) and then cropped in half to obtain twice as many samples.
- **Acoustic guitar:** We manually select 44 monophonic acoustic guitar recordings. The performer played both accompaniment and lead parts, while the performance was captured using either a high-end condenser microphone or a electromagnetic pickup. Clips longer than 19 seconds are trimmed to retain only the initial portion.
- **Electric guitar:** We select 40 stereophonic samples of electric guitar performances, comprising both lead and rhythm guitar parts. The recordings are processed through various (saturating) guitar amplifiers, enriching the harmonic content. Having discarded leading silence, clips longer than 18 seconds are trimmed to retain only the initial portion.
- **Vocals:** We manually select 60 monophonic vocal performances, aiming for an approximate balance between male and female voices, as well as a variety of vocal techniques. Clips longer than 22 seconds are trimmed to retain only the initial portion.

All audio files are natively sampled at 44.1 kHz. Thus, no resampling was applied. Conversely, if needed, audio files are converted from stereo to mono by channel averaging. Every test clip is normalized at -25 LUFS using `pyloudnorm` before being saved in a 16-bit PCM format.

The histogram in Figure 1 illustrates the distribution of sound clip durations across the entire blind test set. Figure 2a, Figure 2b, and Figure 2c, respectively, show the number of clips, the distribution of clip durations, and the number of (simulated) packet losses for each musical instrument.

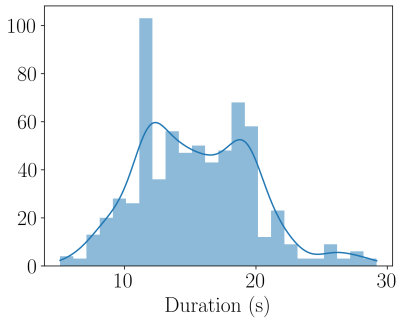


Fig. 1. Duration of audio clips in the 2025 blind test set.

D. Evaluation Procedure

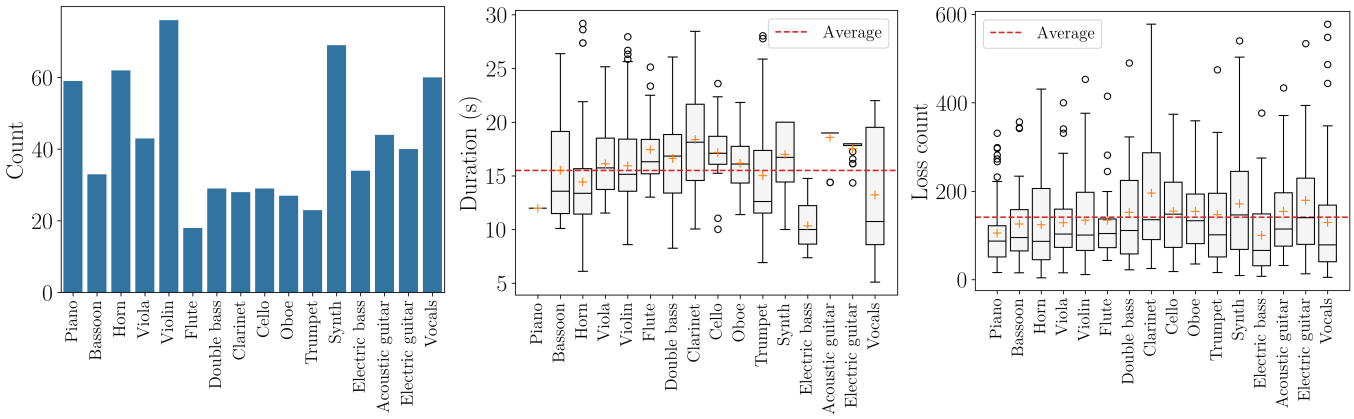
Participants were instructed to download the blind test set, apply their proposed PLC system to all audio clips, and submit the enhanced audio files. As in the 2024 edition, no models were collected or executed during the evaluation phase. System performance was evaluated with a Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) listening test [16]. From the blind test set, ten excerpts were selected to represent a variety of musical instruments. The listening experiment consisted of ten *trials*, where each processed condition was assessed against the clean reference with respect to Basic Audio Quality (BAQ). In each trial, the system achieving the highest *average score* was designated the winner of that trial. The overall team ranking (Table I) was determined by the number of trials won, with the average score across all trials applied as a tie-breaker criterion.

E. Baseline System

For the IEEE-IS^2 2025 Music Packet Loss Concealment Challenge, we adopted PARCnet-IS^2 as the official baseline system. Originally introduced for the 2024 edition where it showcased state-of-the-art performance, PARCnet-IS^2 builds on the original PARCnet architecture proposed in [12].

The PARCnet architecture combines two parallel components: an autoregressive (AR) linear predictor and a feed-forward neural network. The AR model is fitted in real-time applying the autocorrelation method with white noise compensation [17] within a sliding context window. Meanwhile, the neural network learns to predict the residual signal that the finite-memory AR model cannot account for, enhancing the quality of the compound reconstruction.

Several modifications were made to adapt PARCnet [12] for the multi-instrument fullband music PLC task. In particular, PARCnet-IS^2 was trained on Medley-solos-DB [18], which encompass eight musical instruments, rather than the MAESTRO dataset [19], which only includes virtuoso piano recordings. The audio sampling rate was increased from 32 kHz up to 44.1 kHz. Correspondingly, the system was reconfigured to predict packets of 512 samples instead of 320. Training was run for 250,000 steps using an L^1-loss function instead of the



(a) Number of audio clips. (b) Clip duration. (c) Loss count.

Fig. 2. Overview of the 2025 blind test set across musical instruments. Red dashed lines represent global average values; orange crosses indicate the average value for each instrument.

L^2-loss employed in the original paper. The neural network valid context was extended to eight packets, the AR model order was increased to $p = 256$, and the cross-fade length between packets, used to smooth transitions between predicted and valid frames, was raised from 80 to 256 samples. The model also applies cross-fading between packets within a burst loss, and normalizes packets that take values outside the audio range. Full implementation details and pretrained model weights are available online.[4]

III. Team Submissions

We received submissions from three teams. Below, we briefly review the main contributions of each system (in alphabetical order). For further details on the proposed methods, please refer to the respective technical reports [5]–[7].

A. Systems Overview

- **Aironi et al.** [5] submitted a PLC method based on a bin2bin Generative Adversarial Network (GAN) [20], which synthesizes audio conditioned on the output of a order-256 linear predictor. To ensure a smooth inbound transition (left boundary), linear predictive coding (LPC) and neural network estimates are cross-faded. The model employs a convolutional U-Net backbone trained with a combination of spectral convergence, log-magnitude STFT loss, and least-squares conditional GAN objectives. The authors used data from Medley-solos-DB [18], the Good-sounds.org collection [21], and 45 additional hours of audio synthesized from MIDI using soundfonts. MIDI files were sourced from the MAESTRO dataset [19]. Differently from the team's 2024 submission [22], this year's *bin2bin-v2* incorporates depthwise separable convolutions and optimized LPC to reduce memory usage and computational complexity.
- **Daniotti et al.** [6] submitted an extension of the baseline system that involves iteratively training instances of the neural network, and for each iteration an increasing number of packets in the buffer is dropped and concealed with a surrogate model. With the surrogate model being the model trained in the previous iteration, this strategy aims to tackle the increasing divergence of predictions when the past audio buffer contains previously predicted packets, which is common in the case of bursty packet losses. Additionally, the authors employ Temporal Feature-Wise Linear Modulation (TFiLM) [23] as a conditioning strategy to encourage the system to leverage positional information of previously concealed packets and prioritize valid packets. TFiLM uses a recurrent neural network to infer affine transformation parameters from a binary mask of concealed packets within the input buffer, applying these transformations to the activation maps of a PARCnet-IS^2 model.
- **Severi** [7] submitted a sparse linear prediction algorithm based on Orthogonal Matching Pursuit [24]. The PLC method iteratively selects non-contiguous lags, leveraging efficient correlation updates that avoid explicit residual computation. Lost packets are reconstructed autoregressively: the first 20 samples are estimated with an order-8 AR model parameterized using the Burg method [25], while the remaining samples are predicted with a sparse autoregressive (SAR) model of maximum order three. AR and SAR estimates are cross-faded, with the optimal transition point chosen by minimizing the squared error between the predicted signals over a sliding window. The system also implements outbound (right boundary) cross-fading. For further details, we refer the reader to the recent study presenting the approach [26].

B. Real-Time Performance

Challenge participants were required to self-assess the real-time performance of their proposed systems running on

[4]Available: https://github.com/polimi-ispl/2024-music-plc-challenge/tree/main/parcnet-is2

TABLE II
MEAN SQUARED ERROR (↓)

$\times 10^{-3}$	Aironi et al.	Daniotti et al.	Severi	PARCnet-IS^2	Zero-filling
Acoustic guitar	1.296	1.115	1.051	1.001	1.367
Bassoon	0.561	0.506	0.355	0.412	1.353
Cello	0.625	0.568	0.328	0.451	1.204
Clarinet	0.398	0.453	0.387	0.277	1.126
Double bass	1.208	1.230	0.714	1.133	1.642
Electric bass	2.784	1.657	1.460	2.163	1.691
Electric guitar	1.505	1.067	1.211	1.048	1.565
Flute	0.239	0.194	0.287	0.164	0.604
Horn	0.515	0.450	0.344	0.358	1.207
Oboe	0.634	0.390	0.489	0.384	0.877
Piano	1.288	0.903	1.176	0.888	1.351
Synth	1.677	2.335	1.519	1.379	1.603
Trumpet	0.628	0.461	0.514	0.404	0.684
Viola	0.715	0.629	0.533	0.603	1.294
Violin	0.623	0.572	0.619	0.433	0.815
Vocals	0.915	1.645	0.867	0.642	0.936
Average	0.976	0.886	0.741	0.734	1.207

TABLE III
MEAN ABSOLUTE ERROR (↓)

$\times 10^{-3}$	Aironi et al.	Daniotti et al.	Severi	PARCnet-IS^2	Zero-filling
Acoustic guitar	15.773	14.125	12.275	12.934	15.682
Bassoon	9.451	9.296	5.033	7.560	15.909
Cello	10.198	9.935	5.846	8.156	13.142
Clarinet	6.737	8.716	5.669	5.736	14.613
Double bass	15.042	14.579	8.983	13.361	16.038
Electric bass	25.696	18.099	12.921	20.002	16.289
Electric guitar	20.073	16.420	14.621	15.800	19.837
Flute	5.134	5.834	4.574	4.149	9.049
Horn	7.403	8.111	4.736	5.827	13.017
Oboe	9.027	7.989	6.400	6.717	11.727
Piano	17.936	14.421	14.914	13.931	17.168
Synth	20.525	18.855	16.608	17.151	18.747
Trumpet	7.227	8.033	5.499	5.619	7.584
Viola	10.802	10.251	8.043	9.132	14.377
Violin	10.483	10.093	9.256	8.352	11.676
Vocals	12.799	12.564	10.174	10.345	12.960
Average	12.769	11.708	9.097	10.298	14.238

TABLE IV
LOG-SPECTRAL DISTANCE (↓)

	Aironi et al.	Daniotti et al.	Severi	PARCnet-IS^2	Zero-filling
Acoustic guitar	0.250	0.228	0.269	0.227	0.491
Bassoon	0.238	0.307	0.261	0.299	0.711
Cello	0.263	0.262	0.260	0.251	0.586
Clarinet	0.295	0.355	0.374	0.344	0.820
Double bass	0.244	0.259	0.268	0.237	0.669
Electric bass	0.263	0.259	0.279	0.247	0.527
Electric guitar	0.263	0.347	0.302	0.379	0.922
Flute	0.242	0.298	0.304	0.287	0.516
Horn	0.247	0.312	0.284	0.301	0.690
Oboe	0.299	0.382	0.376	0.397	0.749
Piano	0.236	0.278	0.284	0.320	0.761
Synth	0.290	0.355	0.385	0.386	0.798
Trumpet	0.295	0.339	0.347	0.327	0.575
Viola	0.233	0.231	0.245	0.224	0.489
Violin	0.256	0.263	0.309	0.260	0.453
Vocals	0.292	0.281	0.325	0.278	0.503
Average	0.263	0.297	0.305	0.298	0.641

TABLE V
SIGNAL-TO-DISTORTION RATIO (↑)

(dB)	Aironi et al.	Daniotti et al.	Severi	PARCnet-IS^2	Zero-filling
Acoustic guitar	7.762	7.072	9.393	8.757	5.076
Bassoon	11.412	3.616	18.029	11.169	5.178
Cello	8.961	3.961	13.100	9.377	4.982
Clarinet	14.437	2.331	15.866	12.318	4.588
Double bass	6.873	3.183	11.711	6.736	4.890
Electric bass	3.622	4.524	10.631	5.118	5.263
Electric guitar	6.236	5.548	9.099	7.594	4.376
Flute	10.970	−1.261	11.858	9.505	5.476
Horn	12.317	0.565	14.772	11.387	5.530
Oboe	9.939	0.257	11.730	9.172	5.528
Piano	6.476	6.817	7.597	8.124	4.566
Synth	5.855	5.865	7.443	6.780	4.817
Trumpet	8.113	−5.931	8.888	6.105	5.561
Viola	9.733	6.169	11.553	10.351	5.274
Violin	7.856	3.792	8.318	8.490	5.104
Vocals	6.755	4.203	8.554	7.298	4.779
Average	8.582	3.169	11.159	8.642	5.062

TABLE VI
SCALE-INVARIANT SIGNAL-TO-DISTORTION RATIO (↑)

(dB)	Aironi et al.	Daniotti et al.	Severi	PARCnet-IS^2	Zero-filling
Acoustic guitar	4.990	4.928	7.574	6.758	2.644
Bassoon	10.267	7.004	17.296	10.475	2.840
Cello	7.268	4.015	12.336	8.133	2.633
Clarinet	12.860	6.439	14.054	11.466	2.237
Double bass	4.634	1.943	10.481	4.856	2.497
Electric bass	−0.654	2.077	9.013	2.132	2.643
Electric guitar	3.042	3.614	6.977	5.588	1.605
Flute	10.050	4.110	10.672	9.207	3.226
Horn	10.985	4.907	14.090	10.501	3.224
Oboe	8.449	4.625	10.763	8.465	3.161
Piano	3.936	5.207	5.785	6.584	2.139
Synth	2.594	3.408	5.133	4.361	2.184
Trumpet	5.541	−0.511	7.025	4.348	3.114
Viola	8.245	5.790	10.691	9.163	2.955
Violin	5.750	3.457	6.269	6.844	2.599
Vocals	3.792	2.388	6.300	4.834	2.386
Average	6.359	3.963	9.654	7.107	2.631

TABLE VII
CQT-BASED PERCEPTUAL ERROR (↓)

	Aironi et al.	Daniotti et al.	Severi	PARCnet-IS^2	Zero-filling
Acoustic guitar	0.639	1.283	1.039	0.755	1.385
Bassoon	0.774	3.118	0.868	1.063	2.265
Cello	0.552	2.370	0.819	0.736	1.756
Clarinet	1.222	6.655	2.151	2.268	4.549
Double bass	0.443	1.229	0.658	0.563	1.405
Electric bass	0.847	1.261	1.181	1.094	1.562
Electric guitar	0.769	2.071	1.276	0.981	2.009
Flute	0.408	4.155	1.440	1.263	2.245
Horn	0.800	5.085	1.253	1.391	3.051
Oboe	1.005	5.414	1.840	2.086	3.351
Piano	0.462	2.344	1.346	0.696	1.467
Synth	1.086	3.270	2.224	1.739	2.613
Trumpet	1.136	7.886	2.613	2.643	3.157
Viola	0.717	2.952	1.207	1.016	2.074
Violin	0.751	2.757	1.542	1.230	1.888
Vocals	0.897	4.079	1.941	1.533	2.347
Average	0.782	3.496	1.462	1.316	2.320

consumer-grade hardware (Intel Core i5 quad-core 2.4 GHz processor or equivalent). A system was considered real-time if the processing time per packet was shorter than its duration.

- **Aironi et al.** [5] measured execution time on a 2016 Intel Core i7-6850K 3.60 GHz processor. Using four CPU cores, the system processes an audio frame in an average of 8.9 ms, thereby enabling real-time concealment of 11.6 ms gaps.
- **Daniotti et al.** [6] report average TorchScript inference times of 15.85 ms on an Intel Core i9-10940X 3.30 GHz CPU (slower than real-time) and 8.9 ms on an NVIDIA GeForce RTX 4090 GPU with 24 GB of VRAM.
- **Severi** [7] presented a highly optimized C++ implementation that achieves sub-millisecond execution on a single CPU core. On a Raspberry Pi 4 (standard ARM64 architecture), the author reports that model fitting takes approximately 245 µs on average, while sample prediction requires less than 20 ns per sample (10.2 µs per packet).

IV. Evaluation

A. Objective Evaluation

Here, we provide a brief overview of the objective metrics considered as part of the systems evaluation. It is important to emphasize that no objective metric to date has been definitely proved to work for music-oriented PLC algorithms. As such, objective metrics do not play a role in determining the team ranking (Table I), which, instead, was determined solely from the outcome of a MUSHRA-like listening test (Section IV-B).

Time-domain metrics: The blind test set (Section II-C) contains audio files of varying lengths, each subject to a different number of packet losses. To prevent biases in the evaluation, we calculate time-domain metrics over short-time segments spanning five packets centered around each loss, and then average the results.

Let $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i$ be the vectors containing the clean and enhanced segments centered at the i-th packet loss, respectively. We compute the Mean Squared Error (MSE), Mean Absolute Error (MAE), Signal-to-Distortion Ratio (SDR) [27]

$$\text{SDR} := 10 \log_{10} \frac{\|\mathbf{y}_i\|^2}{\|\mathbf{y}_i - \hat{\mathbf{y}}_i\|^2}, \tag{1}$$

and Scale-Invariant Signal-to-Distortion Ratio (SI-SDR) [27]

$$\text{SI-SDR} := 10 \log_{10} \frac{\|\alpha\mathbf{y}_i\|^2}{\|\alpha\mathbf{y}_i - \hat{\mathbf{y}}_i\|^2}, \tag{2}$$

where $\alpha = \langle \hat{\mathbf{y}}_i, \mathbf{y}_i \rangle / \|\mathbf{y}_i\|^2$, which reduces to (1) with $\alpha = 1$.

Spectral metrics: Differently from time-domain metrics, spectral metrics are computed on the entire audio signals.

We report the Log-Spectral Distance (LSD) [28]

$$\text{LSD} = \frac{1}{M} \sum_{m=0}^{M-1} \sqrt{\frac{1}{K} \sum_{k=0}^{K} \log|Y[m,k]|^2 - \log|\hat{Y}[m,k]|^2}\,, \tag{3}$$

where $Y[m,k]$ and $\hat{Y}[m,k]$ denote the short-time Fourier transform (STFT) of the reference and enhanced signal, $y[n]$ and $\hat{y}[n]$, respectively. The STFT with M frames and $(K+1)$ frequency bins is computed using a 2048-sample Hann window with 75% overlap.

Moreover, inspired by [29], we adapt a recent perceptual metric leveraging Constant-Q Transform (CQT) representations [30]. For each enhanced signal, we take the dB-scale normalized complex CQT error magnitude relative to the original signal. The resulting values are then averaged across time and frequency to yield a single scalar measure. Although [29] focuses on assessing glitch audibility rather than evaluating music-oriented PLC algorithms, we choose the perceptually-motivated CQT parameters proposed by the authors: 48 bins per octave, a minimum frequency of 10 Hz, and a minimum window length of 2048 samples.

B. Subjective Evaluation

As in the previous edition, the IEEE-IS^2 2025 Music Packet Loss Concealment Challenge ranking was determined through a MUSHRA-like test. To provide a fair test bench, ten signals from the blind test set were handpicked by experts who only had access to the lossy audio files (zero-filling). Overall, one sample was selected from each of the following classes: piano, strings (violin), woodwinds (clarinet), brass instruments (trumpet), electric bass, acoustic guitar, electric guitar, singing voice, synth pad, and synth lead.

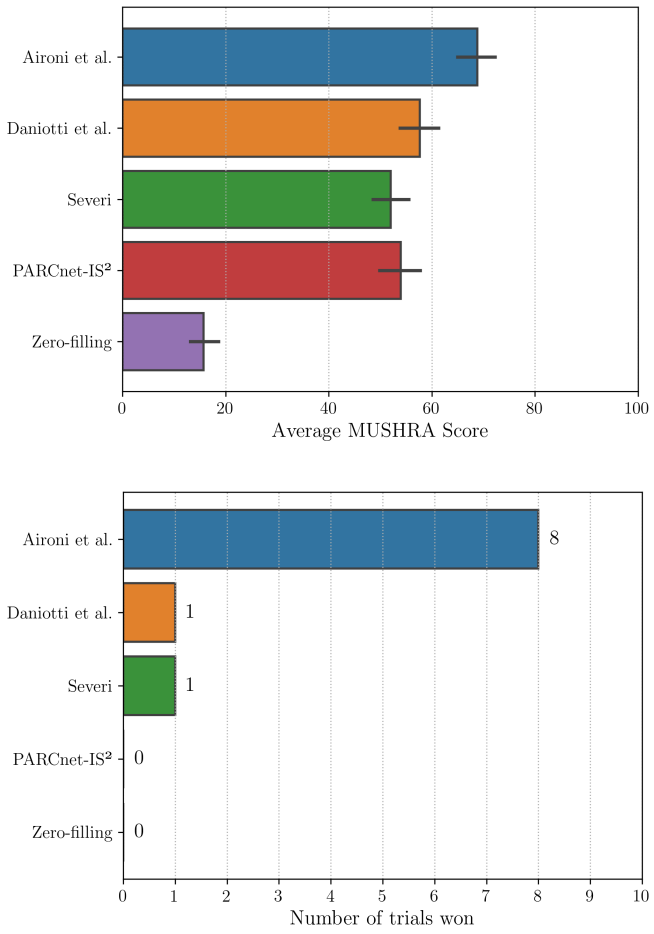
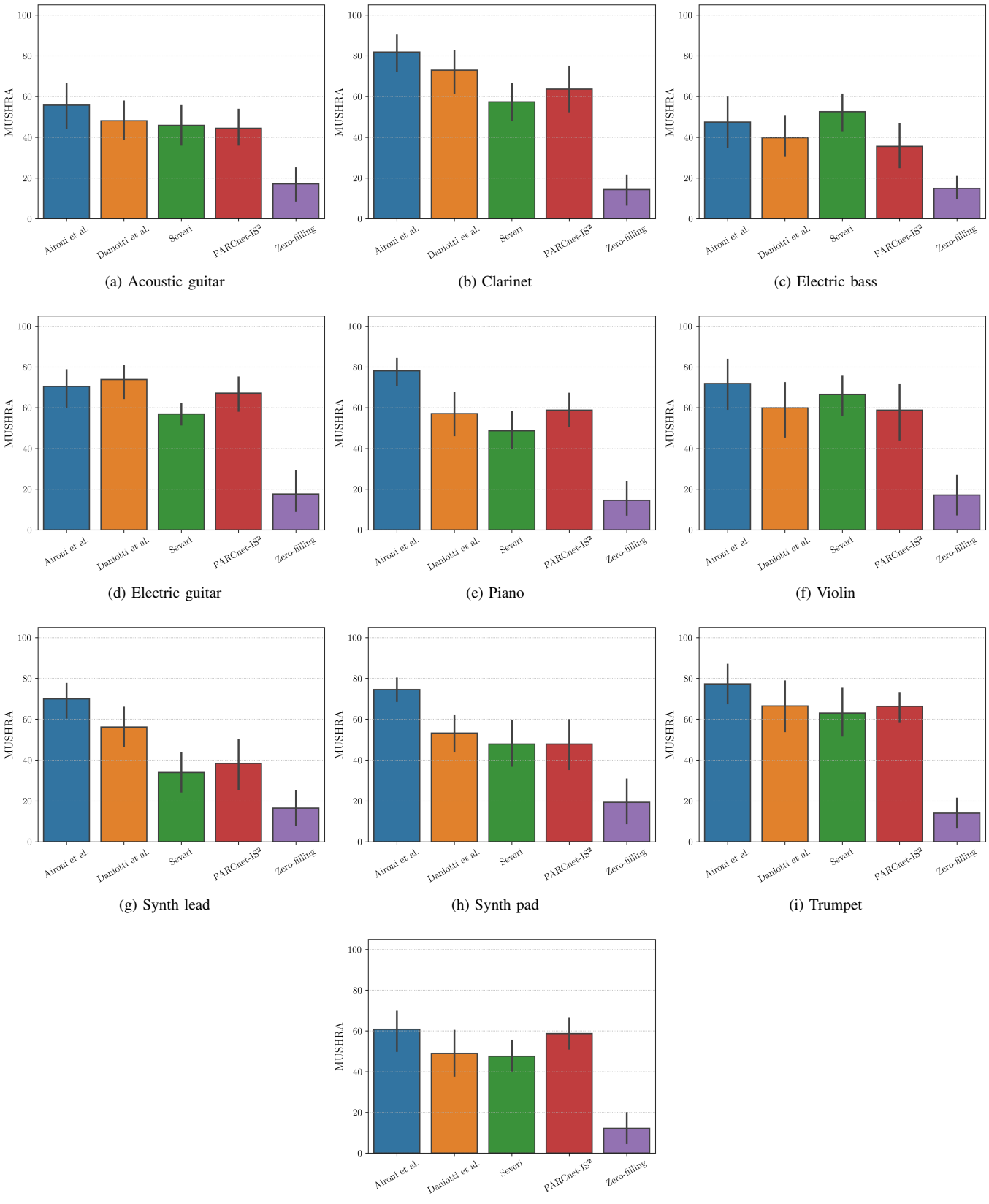


Fig. 3. Results of the MUSHRA test.

The test was conducted using webMUSHRA [31], a state-of-the-art Web Audio API-based software platform compliant with Rec. ITU-R BS.1534-3 [16].

For each of the ten excerpts, the (undisclosed) clean audio file was used as Reference, whereas the clip degraded with zero-filling was considered as the Anchor. After an initial training page where four held-out pairs of clean and zero-filling audio examples were presented (violin, opera singing, steel-string acoustic guitar, and piano), participants were asked to rate the similarity of each test condition with the Reference on a scale of 0 to 100. On each page, six conditions were assessed, including the output of the baseline system, the Hidden Reference, and the Anchor. The names of the test conditions were hidden, and the order of both test items and trials was randomized. Volume adjustments were only allowed during the training phase. Then, subjects were asked to keep the level constant for the duration of the test.



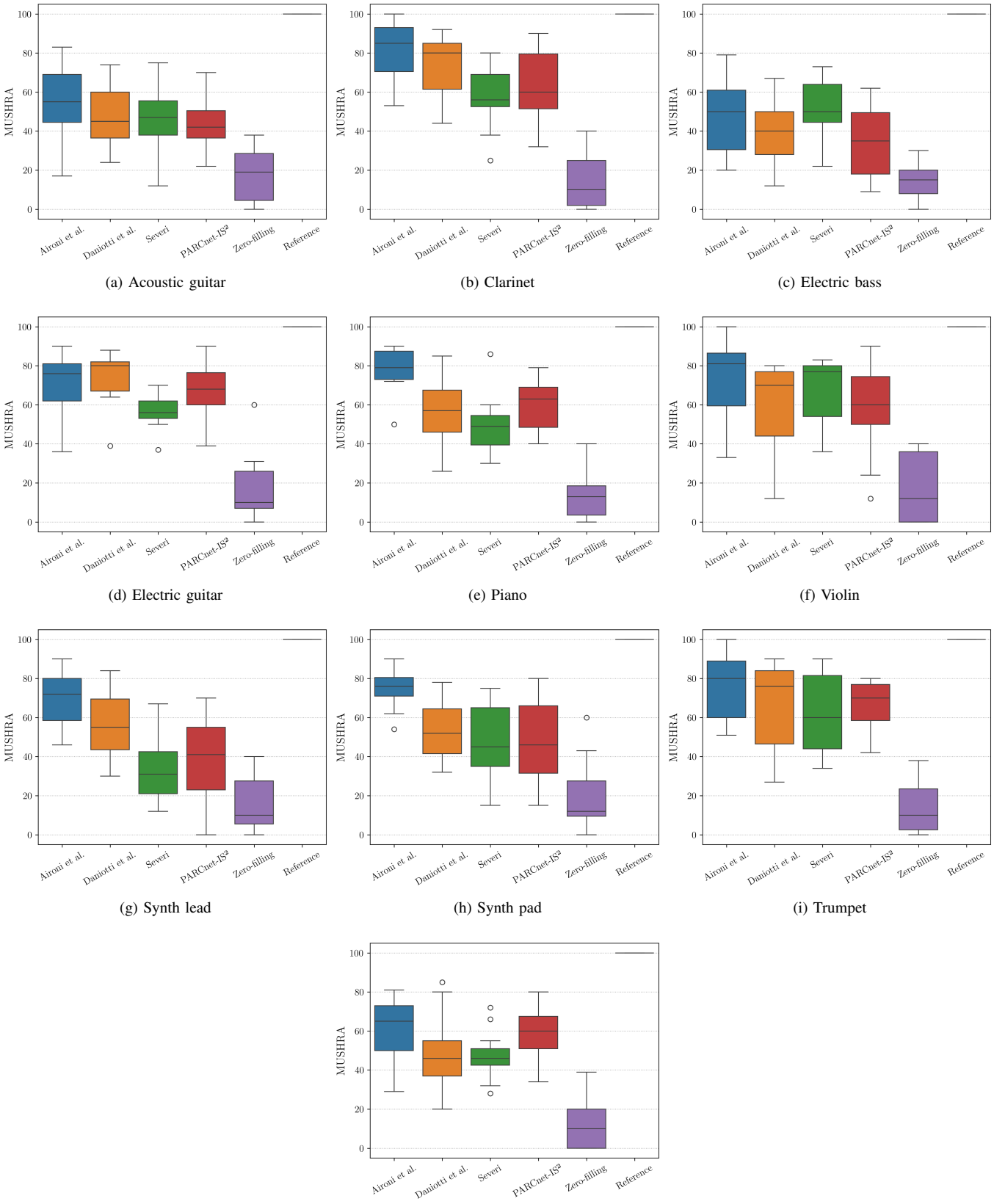
(a) Acoustic guitar (b) Clarinet (c) Electric bass

(d) Electric guitar (e) Piano (f) Violin

(g) Synth lead (h) Synth pad (i) Trumpet

(j) Vocals

Fig. 4. Average scores and 95% confidence intervals of the individual trials in the MUSHRA test.



(a) Acoustic guitar (b) Clarinet (c) Electric bass

(d) Electric guitar (e) Piano (f) Violin

(g) Synth lead (h) Synth pad (i) Trumpet

(j) Vocals

Fig. 5. Box-and-whisker plots of the individual trials in the MUSHRA test.

Thirteen experienced assessors, aged 25–50 (average: 34.9), took part in the listening test. In accordance with the ITU-R BS.1534-3 post-screening criteria [16], two assessors were excluded from the aggregated responses because they rated the Hidden Reference below 90 in more than 15% of test items (two out of ten trials). The remaining participants completed the task in under 40 minutes, reported no hearing impairments, and had an average of 9.9 years of musical training (SD: 8.67). Six had prior experience with NMP. Assessors were affiliated with the Image and Sound Processing Lab (ISPL) at Politecnico di Milano, the RADIOLABS research consortium, or were music professionals.

V. Results

Objective evaluation: Tables II through VII present the objective metrics described in Section IV-A. The best values for each instrument are underlined. On average, PARCnet-IS^2 has the lowest MSE, Severi performs best on the remaining time-domain metrics (MAE, SDR, SI-SDR), while Aironi et al. achieve the lowest scores in both spectral metrics.

Subjective evaluation: Figure 3 shows the average scores and 95% confidence intervals obtained across all trials in the MUSHRA test (top) and the number of trials won by each PLC method (bottom). Figure 4 and Figure 5 show the average scores and the box-and-whisker plots of every trial in the listening test, respectively. The results of the MUSHRA test determined the final ranking shown in Table I.

Team ranking: Aironi et al. [5] achieved first place, winning eight out of ten trials. Neither the baseline model nor zero-filling secured any wins. Daniotti et al. [6] and Severi [7] each won a single trial (Electric guitar and Electric bass, respectively). However, Daniotti et al. [6] obtained a higher average MUSHRA score across all trials (57.66 ± 20.30) compared to Severi [7] (52.01 ± 18.24), thereby breaking the tie and ranking second overall, with Severi [7] placed third.

Final remarks: The challenge results indicate that strong performance on time-domain metrics does not necessarily translate to perceptual preference. This outcome was largely anticipated, and the subjective evaluation appears to confirm the hypothesis. In contrast, spectral metrics align much more closely with the MUSHRA results, with Aironi et al. [5] ranking first in both objective and subjective tests.

VI. Conclusions

In this paper, we outlined the setup and outcomes of the IEEE-IS^2 2025 Music Packet Loss Concealment Challenge. Three teams submitted systems exploring different PLC paradigms, spanning from classical signal processing to advanced neural network architecture. We introduced a diverse blind test set covering 16 instrument classes, designed to evaluate high-fidelity multi-instrument PLC systems under adverse packet loss conditions. The evaluation employed both objective metrics and a subjective MUSHRA-like listening test, which ultimately determined the final team ranking. While objective metrics provide useful insights into reconstruction accuracy, the subjective evaluation remains critical to assess perceived audio quality, especially in musical applications.

Overall, the challenge highlighted the effectiveness of combining linear predictive coding and neural architectures, as well as the potential of adversarial generative models for high-quality music-oriented PLC. The findings from this edition provide a foundation for future research toward robust and perceptually transparent packet loss concealment in diverse musical scenarios.

Acknowledgment

The authors gratefully acknowledge the expert assessors for their valuable participation in the listening test.

References

[1] L. Turchet, M. Lagrange, C. Rottondi, G. Fazekas, N. Peters, J. Østergaard, F. Font, T. Bäckström, and C. Fischione, "The internet of sounds: Convergent trends, insights, and future directions," *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11 264–11 292, 2023.

[2] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.

[3] A. Carôt and C. Werner, "Fundamentals and principles of musical telepresence," *Journal of Science and Technology of the Arts*, vol. 1, no. 1, pp. 26–37, 2009.

[4] P. Verma, A. I. Mezza, C. Chafe, and C. Rottondi, "A deep learning approach for low-latency packet loss concealment of audio signals in networked music performance applications," in *2020 27th Conference of Open Innovations Association (FRUCT)*, 2020, pp. 268–275.

[5] C. Aironi, L. Gabrielli, S. Cornell, and S. Squartini, "Restoring music integrity via predictive modeling and spectral inpainting," Università Politecnica delle Marche & Carnegie Mellon University, Tech. Rep., 2025. [Online]. Available: https://internetofsounds.net/public_downloads/Aironi_tech_report_2025.pdf

[6] F. Daniotti and L. Turchet, "Conditioning PARCnet with TFiLM for robust packet loss concealment," University of Trento, Tech. Rep., 2025. [Online]. Available: https://internetofsounds.net/public_downloads/Daniotti_tech_report_2025.pdf

[7] L. Severi, "Orthogonal matching pursuit based linear prediction for real-time packet loss concealment," Politecnico di Torino, Tech. Rep., 2025. [Online]. Available: https://internetofsounds.net/public_downloads/Severi_tech_report_2025.pdf

[8] "A high quality low-complexity algorithm for packet loss concealment with G.711," Rec. ITU-T G.711 Appendix I, International Telecommunications Union, Geneva, Switzerland, Sep. 1999.

[9] L. Diener, S. Sootla, S. Branets, A. Saabas, R. Aichner, and R. Cutler, "INTERSPEECH 2022 audio deep packet loss concealment challenge." in *Proc. Interspeech 2022*, 2022, pp. 580–584.

[10] L. Diener, S. Branets, A. Saabas, and R. Cutler, "The ICASSP 2024 audio deep packet loss concealment grand challenge," *IEEE Open Journal of Signal Processing*, vol. 6, pp. 231–237, 2025.

[11] M. Sacchetto, Y. Huang, A. Bianco, and C. Rottondi, "Using autoregressive models for real-time packet loss concealment in networked music performance applications," in *Proc. of the 17th International Audio Mostly Conference*, 2022, pp. 203–210.

[12] A. I. Mezza, M. Amerena, A. Bernardini, and A. Sarti, "Hybrid packet loss concealment for real-time networked music applications," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 266–273, 2024.

[13] A. I. Mezza and A. Bernardini, "The IEEE-IS² 2024 music packet loss concealment challenge," *arXiv preprint arXiv:2409.18564*, 2024.

[14] D. Thery and B. F. Katz, "Anechoic audio and 3D-video content database of small ensemble performances for virtual concerts," in *Proc. of the 23rd International Congress on Acoustics (ICA)*, 2019, pp. 739–746.

[15] C. J. Steinmetz and J. Reiss, "pyloudnorm: A simple yet flexible loudness meter in Python," in *Audio Engineering Society Convention 150*. Audio Engineering Society, 2021.

[16] "Method for the subjective assessment of intermediate quality level of audio systems," Rec. ITU-R BS.1534-3, International Telecommunications Union, Geneva, Switzerland, Jun. 2021.

[17] P. Kabal, "Ill-conditioning and bandwidth expansion in linear prediction of speech," in *Proc. of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, 2003, pp. 824–827.

[18] V. Lostanlen, C.-E. Cella, R. Bittner, and S. Essid, "Medley-solos-DB: a cross-collection dataset for musical instrument recognition," Zenodo, Sep. 29, 2019. doi: 10.5281/zenodo.3464194.

[19] C. Hawthorne, A. Stasyuk, A. Roberts, I. Simon, C.-Z. A. Huang, S. Dieleman, E. Elsen, J. Engel, and D. Eck, "Enabling factorized piano music modeling and generation with the MAESTRO dataset," in *International Conference on Learning Representations*, 2019.

[20] C. Aironi, S. Cornell, L. Serafini, and S. Squartini, "A time-frequency generative adversarial based method for audio packet loss concealment," in *2023 31st European Signal Processing Conference (EUSIPCO)*, 2023, pp. 121–125.

[21] G. Bandiera, O. Romani Picas, H. Tokuda, W. Hariya, O. Koji, and X. Serra, "Good-sounds.org: A framework to explore goodness in instrumental sounds," in *Proc. of the 17th International Society for Music Information Retrieval Conference (ISMIR)*, 2016, pp. 414–419.

[22] C. Aironi, L. Gabrielli, S. Cornell, and S. Squartini, "Enhancing music packet loss concealment with generative spectrogram inpainting," Università Politecnica delle Marche & Carnegie Mellon University, Tech. Rep., 2024. [Online]. Available: https://internetofsounds.net/public_downloads/Aironi_tech_report.pdf

[23] S. Birnbaum, V. Kuleshov, Z. Enam, P. W. W. Koh, and S. Ermon, "Temporal FiLM: Capturing long-range sequence dependencies with feature-wise modulations." *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[24] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. of 27th Asilomar conference on signals, systems and computers*. IEEE, 1993, pp. 40–44.

[25] K. Vos, "A fast implementation of Burg's method," *OPUS codec*, 2013.

[26] L. Severi and C. Rottondi, "Sparse linear prediction for packet loss concealment in networked music performances," in *Proc. of the 6th International Symposium on the Internet of Sounds (IS2 2025)*, 2025.

[27] J. Le Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, "SDR—half-baked or well done?" in *ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 626–630.

[28] A. Gray and J. Markel, "Distance measures for speech processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 5, pp. 380–391, 1976.

[29] L. Vignati and L. Turchet, "On the lack of a perceptually-motivated evaluation metric for packet loss concealment in networked music performances," Unpublished.

[30] C. Schörkhuber and A. Klapuri, "Constant-Q transform toolbox for music processing," in *7th Sound and Music Computing Conference*, 2010, pp. 3–64.

[31] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, and J. Herre, "webMUSHRA — A comprehensive framework for web-based listening tests," *Journal of Open Research Software*, vol. 6, 2018.