

Tweaking autoregressive methods for inpainting of gaps in audio signals

Ondřej Mokřý

Department of Telecommunications
Brno University of Technology
Czech Republic

Pavel Rajmic

Department of Telecommunications
Brno University of Technology
Czech Republic

Abstract—A novel variant of the Janssen method for audio inpainting is presented and compared to other popular audio inpainting methods based on autoregressive (AR) modeling. Both conceptual differences and practical implications are discussed. The experiments demonstrate the importance of the choice of the AR model estimator, window/context length, and model order. The results show the superiority of the proposed gap-wise Janssen approach using objective metrics, which is confirmed by a listening test.

Index Terms—audio, autoregression, inpainting, interpolation, comparison, packet loss concealment

I. INTRODUCTION

Audio inpainting is a challenging signal processing task, where missing parts of an audio signal have to be completed. For a human listener, the result should be as pleasant as possible and ideally free of artifacts. Previously proposed audio inpainting solutions cover a wide range of approaches, from autoregressive modeling [1], [2], [3], [4], through optimization methods [5], [6], [7], [8], [9], heuristics [10], graph-based methods [11] to deep learning [12], [13], [14], [15] and hybrid approaches [16].

For signal gaps of up to ca 80 milliseconds, the iterative method of Janssen et al. [1] proposed in 1986 constantly ranks among the best, according to numerous studies [6], [9], [17], [18]. The extrapolation methods [3], [4], [19], [20] are non-iterative and utilize a twofold extrapolation (from left to right and right to left) while the two particular solutions are blended together using a crossfading scheme. Such an approach belongs to the most popular, perhaps due to its simplicity and speed, and is actually used in the Matlab function `fillgaps`. The patented method of Etter [2] considers the just mentioned approach suboptimal and proposes to aggregate the two extrapolation directions in a single optimization criterion.

Besides slight variances in how to *model the signal*, different algorithms for *estimation of the coefficients* are also available [21], which will be discussed in detail further on.

In this paper, we review the principle of autoregression-based methods, point out the main differences between the particular popular approaches, and present computational experiments on two audio datasets. Most importantly, we propose a new variant of the Janssen algorithm [1] not present in the

literature and examine its performance compared with known approaches. This novel method does not rely on frame-wise signal processing, in contrast to the original method.

This paper considers only gaps up to 80 ms in length. For larger gap sizes, autoregression usually starts to become inefficient. A number of non-autoregressive methods cited above were designed to cope with larger gaps; however, note that it is challenging to compare against such methods for at least two reasons: first, they are typically trained on a specific class of data, while AR modeling is data-independent; second, these methods fill the gaps with material which may be pleasant to listen to, but it does not have to align with the original audio, making it hard to objectively judge the reconstruction quality.

II. MODELING AUDIO AS AN AUTOREGRESSIVE PROCESS

A signal $\mathbf{x} = [x_1, \dots, x_N]^T \in \mathbb{R}^N$ is said to be modeled as an autoregressive (AR) process, if it satisfies

$$\sum_{i=1}^{p+1} a_i x_{n+1-i} = e_n, \quad n = 1, \dots, N + p, \quad (1)$$

where $\mathbf{e} = [e_1, \dots, e_{N+p}]^T$ is a realization of a zero-mean white noise process, and $\mathbf{a} = [1, a_2, \dots, a_{p+1}]^T$ are referred to as the AR coefficients [21, Def. 3.1.2]. The order p defines the range of indexes that determine the output in the current time instance. As such, the frequency resolution increases with increasing p . The above is closely related to the notion of convolution, and it is actually in accordance with the convention of the `lpc` function in Matlab.

From the perspective of model fitting, the vector $\mathbf{e} \in \mathbb{R}^{N+p}$ is called the *residual error* [22, Sec. 8.2.2]. Given the observed signal \mathbf{x} and the order p , the coefficients of the AR model are usually estimated via the optimization problem

$$\arg \min_{\mathbf{a}} \frac{1}{2} \|\mathbf{e}(\mathbf{a}, \mathbf{x})\|^2, \quad (2)$$

where the error $\mathbf{e}(\mathbf{a}, \mathbf{x})$, defined by its entries in (1), is a function of both the signal \mathbf{x} and the coefficients \mathbf{a} . Problem (2) can be effectively solved using an estimate of the autocorrelation and the Levinson–Durbin algorithm [23], [24]. Such an approach is typically referred to as the LPC for its connection to the linear prediction coefficients.

In contrast to the LPC, the Burg algorithm [25] for the estimation of the AR parameters involves an extra assumption

The work was supported by the Czech Science Foundation (GAČR) Project No. 23-07294S.

that the *same* parameters should model both the signal \mathbf{x} and its version flipped in time. In effect, this results in another quadratic term extending (2). Kauppinen and Roth prefer the Burg algorithm for audio signal extrapolation [19, Sec. 4.2] since the underlying all-pole filter obtained this way is stable [26, Sec. 12.3.3]. A frequency-warped Burg algorithm has been proposed [20] that allows focusing on specified spectral bands; however, the effect of warping is comparable to increasing the model order p in the non-warped case [20, Sec. 5].

Additional terms can be optionally appended to the objective (2) that *regularizes* either the reconstructed signal or the AR coefficients, as proposed in [27], [28].

III. AUDIO INPAINTING USING AUTOREGRESSION

To formalize the problem of audio inpainting, assume that the observed signal $\mathbf{x} \in \mathbb{R}^N$ consists of reliable samples identified by the set of indices $M \subset \{1, \dots, N\}$, and vacant samples at positions $\bar{M} = \{1, \dots, N\} \setminus M$. The goal of inpainting is to estimate the missing samples at the positions \bar{M} . In the so-called consistent case, the samples at positions M are meant to be preserved, i.e., any candidate solution $\hat{\mathbf{x}}$ to the inpainting problem should satisfy $\hat{x}_n = x_n$ for all $n \in M$.

In this work, the focus is on the practical scenario where the signal contains gaps, i.e., segments of consecutive lost samples, surrounded by an intact context, as found in packet loss concealment, for instance [16]. For such a scenario, two approaches based on AR modeling are applicable.

First, the extrapolation-based method fits two independent sets of AR parameters for each gap, one for the left-hand context and one for the right-hand context of the gap. These coefficients are then used to extrapolate (predict) both contexts inside the gap, and the forward- and backward-extrapolated signals are then cross-faded. Numerous fading options are possible (see, for instance, [3, Sec. 4.2]); we resort to the raised cosine function used in [2].

Second, the Janssen method operates independently on individual signal frames, typically obtained by windowing with overlaps. In each frame, the iterative Janssen method alternates between the estimation of the AR model for the frame (the current estimate of the missing samples being fixed) and the missing samples in this frame (with a fixed estimate of the model parameters [1]). As such, a problem similar to (2) is solved, where both \mathbf{a} and \mathbf{x} are variables, and \mathbf{x} is constrained to stick to the reliable part of a signal frame. Using the overlap-add procedure, the individually processed frames are joined to form the output.

Finally, we propose a novel use of the Janssen method, which, instead of overlapping frames, treats each gap in the signal separately (hence the name gap-wise Janssen). Here, a single AR model is simultaneously fitted to *both* left- and right-hand contexts of the gap¹. Such a way of using the context is similar to the extrapolation-based method, but the AR coefficients (shared by both contexts) are estimated as in

¹If, furthermore, the Burg algorithm is used to estimate the AR coefficients, it means that a single AR model is assumed not only for the whole gap context, but also for its flipped version.

the Janssen method. Note that neither the frame-wise or the proposed gap-wise Janssen approach is limited to the selected scenario where a *compact* gap is surrounded by reliable context. This is, however, not the case of the extrapolation method, and therefore the experiments stick to that use case.

IV. EXPERIMENTS & RESULTS

To simulate degradation, we consider gap lengths from 10 ms up to 80 ms, and create 10 gaps in each signal at pseudorandom locations.² From AR-based methods, we use all three aforementioned approaches: The extrapolation-based method and the gap-wise Janssen method are applied with a fixed context length of 4096 samples (approx. 93 ms) on each side of the gap (i.e., 8192 samples in total). The frame-wise Janssen uses a frame length of 4096 samples and two window shapes: rectangular and Hann (see, e.g., [30, Sec. V]). All methods are applied with the varying model order p and either the Burg or the LPC algorithm to fit the AR model.

The first measure of reconstruction quality is the signal-to-distortion ratio (SDR). For the reference (i.e., undegraded) signal \mathbf{y} and the reconstruction $\hat{\mathbf{x}}$, SDR in decibels is computed as $\text{SDR}(\mathbf{y}, \hat{\mathbf{x}}) = 10 \log_{10} \frac{\|\mathbf{y}\|^2}{\|\mathbf{y} - \hat{\mathbf{x}}\|^2}$. In our case, the SDR is only computed in the inpainted sections of the signal. The perceived quality of the signal is assessed using PEMO-Q [31], an objective metric which predicts the subjective difference of \mathbf{y} and $\hat{\mathbf{x}}$ in terms of the objective difference grade (ODG), ranging from -4 (very annoying) to 0 (imperceptible). An alternative choice, which is common in other audio processing fields, is PEAQ [32], [33]. However, this metric is not decisive enough in the case of gap inpainting [34].

Our first dataset consists of 9 recordings of individual musical instruments³, taken from the EBU SQAM database [29]. They are sampled at 44.1 kHz and cut to a length of around 7 seconds. Note that further experiments were performed using a longer window/context length of 8192 samples, and also using a mid-scale dataset based upon the music IRMAS database [35], [36]; those results are presented later in Sec. IV-D.

A. Effect of the estimator

First, we evaluate the performance of the inpainting methods depending on the estimator of the AR model parameters. Hence, for each test instance, two versions of each inpainting method are run, using either the LPC or the Burg algorithm.

The results are presented in Fig. 1. The overall distribution of the results in the scatter plots indicates that, in terms of the ODG, the Burg algorithm is clearly favorable in the case of the extrapolation-based inpainting. The preferability of the Burg algorithm is also indicated in the gap-wise Janssen algorithm. Notably, a conclusion in the case of the frame-wise Janssen algorithm depends on the model order and the selected window. With the Hann window, inpainting results

²Signals with fixed masks of the reliable samples are available in the repository <https://github.com/ondrejmkry/TestSignals>.

³We chose solo instruments since AR models are expected to perform well on them; a sum of multiple AR processes generally may not be an AR process of order reasonably comparable with the individual signal orders.

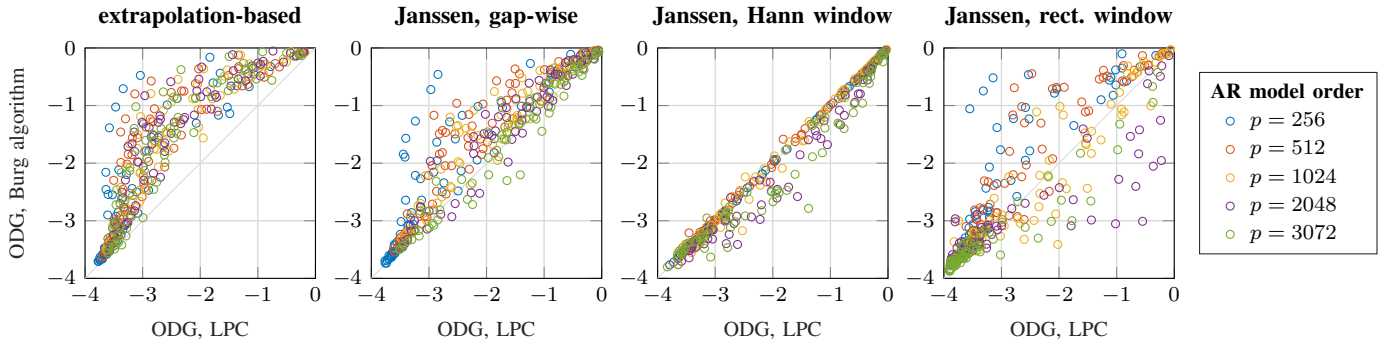


Fig. 1. Comparison of the estimators in terms of PEMO-Q ODG for window/context length 4096 samples. Per each inpainting method, the scatter plot shows the individual results using LPC vs. the Burg algorithm to estimate the AR coefficients. The effect of the model order p is analyzed separately in Fig. 2.

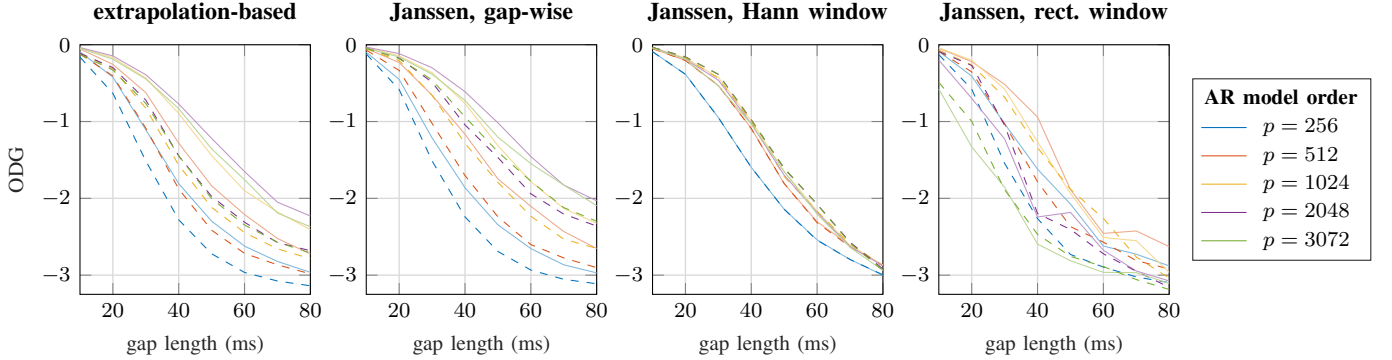


Fig. 2. Comparison of the model order choices in terms of PEMO-Q ODG for window/context length 4096 samples. Per each inpainting method, the plot shows averaged results using LPC (darker shade, dashed line) vs. the Burg algorithm (lighter shade, solid line) to estimate the AR coefficients.

appear to depend on the chosen estimator only for large model orders (LPC scores better in such cases). With the rectangular window, the dependence on the model order is clearly amplified. Furthermore, a large portion of the results in this case implies that the Burg algorithm is a better choice if the model order is low, and vice versa. The differences have been verified as statistically significant using the Wilcoxon signed rank test [37].

Note that in terms of SDR, the differences are less pronounced compared to the ODG, but the conclusions are analogous. The complete results are available at the accompanying webpage.⁴

B. Effect of the model order

The AR model order p plays a crucial role in the modeling and also significantly affects the results, as demonstrated in Fig. 2. Note that Fig. 2 plots the same data as Fig. 1, but with the focus on the effect of the model order and the gap length.

Similarly to the study of preferences between LPC and the Burg algorithm, the model order affects the results differently for different inpainting methods. A clear scheme is observed in the case of the extrapolation-based and the gap-wise Janssen methods, where increasing the model order up to $p = 2048$ results in both a higher SDR and a higher ODG. However, the order $p = 3072$ does not further increase the resulting quality. Note that the same observation holds for the experiment with window/context length 8192, including model order $p = 4096$;

see the supplementary material online. For the frame-wise Janssen, the results are further affected by the chosen window shape: the best results are achieved using $p = 512$ with the rectangular window, and $p = 1024$ with the Hann window.

Furthermore, Fig. 2 reveals that the observed phenomena do not depend on the length of the gap.

C. Comparison with other methods

To provide a context for the results of AR-based methods, selected variants are compared with the methods that belong to the state-of-the-art in optimization-based audio inpainting, namely A-SPAIN [6] and A-SPAIN-MOD [17], applied with the same window length of 4096 samples. The results in terms of SDR and ODG are presented in Fig. 3. The most significant observation is the dominance of the extrapolation-based and, especially, the proposed gap-wise Janssen methods, in particular in gaps larger than 50 ms.

In addition to objective metrics, a subjective listening test was performed on a subset of three audio excerpts (the violin, piano, and clarinet) with gaps of 20, 50 and 80 milliseconds in length, making 9 test signals altogether. Such a combination of signals, gap lengths and reconstruction methods yields a total length of 8.5 minutes of netto audio to evaluate. We ran a MUSHRA-type test [38], using the webMUSHRA environment [39], in which the participants evaluated the quality of reconstructions. The test was run in a quiet music studio, using a professional sound card and headphones. The conditions were identical for all the participants. Ten participants passed the pre- and postscreening selection. We used the signal with

⁴<https://ondrejmkry.github.io/InpaintingAutoregressive/>

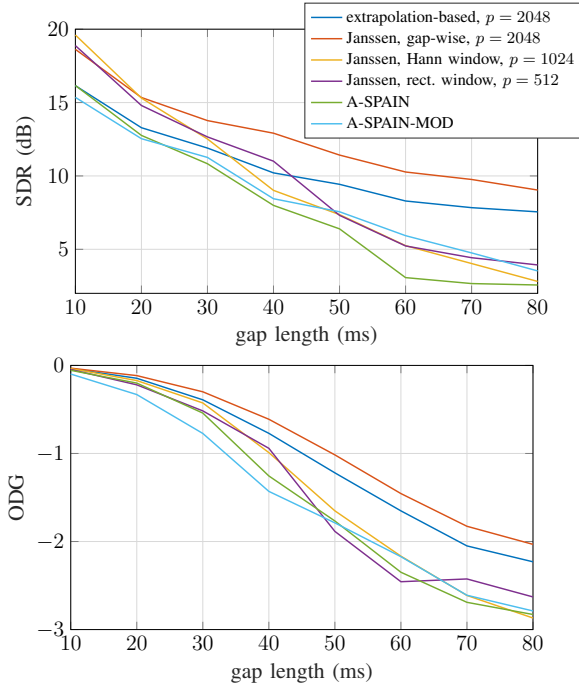


Fig. 3. Comparison of the AR-based methods with SPAIN in terms of SDR (top) and ODG (bottom), averaged over all signals. In this experiment, all AR-based methods used the Burg algorithm to estimate the coefficients, using the best performing order p according to the results reported in Fig. 2.

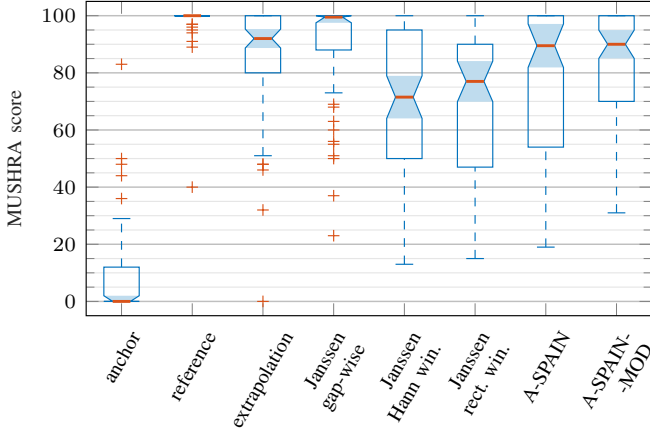


Fig. 4. A boxplot showing the distribution of scores in the listening test. The proposed gap-wise Janssen method proves to be the best performing method, which is also confirmed statistically, since non-overlapping notches (filled areas) imply the difference of medians at the 5% significance level.

gaps as the anchor. The results, summarized in Fig. 4, confirm the alignment of the subjective scores with the ODG metric, presented in Fig. 3, especially the ranking of the extrapolation-based and gap-wise Janssen methods.

D. Testing on a mid-scale dataset, including increased window/context length

An additional, mid-scale comparison was performed, using 60 signals selected from the IRMAS database [35], [36]. Our selection contains a wide range of music characters and genres but avoids recordings with highly pronounced drums. Each

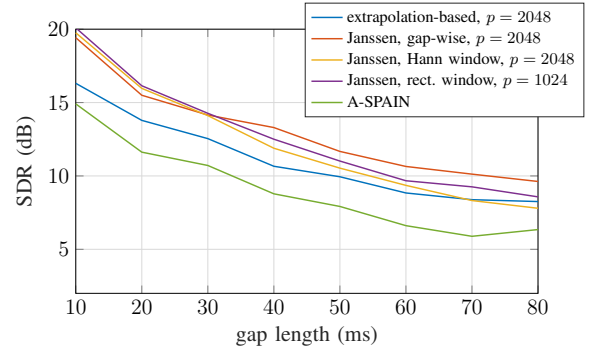


Fig. 5. Comparison of the AR-based methods with SPAIN in terms of SDR, using an increased window/context length of 8192 samples on the solo-instrument dataset. All AR-based methods used the Burg algorithm to estimate the coefficients and the orders were selected as the best-performing for this window/context length. A-SPAIN-MOD is omitted for computational reasons (taking around 3.5 hours per signal).

final excerpt is 7 seconds long. The test material is much more complex than in the solo-instrument case, thus weakening the assumption about the autoregressive nature of signals, which explains obtaining lower values of both SDR and ODG for all the methods considered. However, the results exhibit the same behavior as the small-scale experiment from Fig. 3; in particular, the ranking of the methods was identical.

Finally, all the experiments were repeated using a longer window/context length of 8192 samples. This change is beneficial especially for the windowed methods, see Fig. 5; however, the gap-wise Janssen remains superior on average. For complete results, see the accompanying webpage.

V. CONCLUSION

A gap-wise modification of the classic Janssen method was proposed and evaluated against popular audio inpainting methods based on autoregressive modeling. The experiments demonstrated the importance of the choice of the AR model estimator (i.e., choosing the LPC or the Burg algorithm) and the model order. The concluding tests, both objective and subjective, revealed that the gap-wise Janssen method (using the Burg algorithm) is recommended as an autoregressive reference for inpainting of gaps up to 80 ms; this holds even in comparison with sparsity-based approaches.

If computational speed is an important criterion, note that for all approaches, the computational load is proportional both to the order of the AR model and to the gap length. Moreover, the Burg algorithm is more demanding compared to the LPC. From the perspective of computational time⁵, the extrapolation-based approach is clearly preferable (elapsed times are up to around 0.15 s per signal with $p = 2048$, while the gap-wise Janssen reaches up to 11.5 s per signal with $p = 1024$, and up to 16 s with $p = 2048$; both using the Burg algorithm).

The Matlab codes for the methods discussed in the present paper are publicly available.⁶

⁵The computations were performed on a desktop computer with Intel Core i7-10700K CPU at 3.80GHz with 32 GB RAM.

⁶<https://github.com/ondrejmkry/InpaintingAutoregressive>

REFERENCES

- [1] A. J. E. M. Janssen, R. N. J. Veldhuis, and L. B. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 34, no. 2, pp. 317–330, Apr. 1986.
- [2] W. Etter, "Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1124–1135, 1996.
- [3] Ismo Kauppinen and Jyrki Kauppinen, "Reconstruction method for missing or damaged long portions in audio signal," *Journal of the Audio Engineering Society*, vol. 50, no. 7/8, pp. 594–602, 2002.
- [4] Paulo Antonio Andrade Esquef, Vesa Välimäki, Kari Roth, and Ismo Kauppinen, "Interpolation of long gaps in audio signals using the warped Burg's method," in *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx-03)*, London, UK, Sept. 2003, pp. 18–23.
- [5] Amir Adler, Valentin Emiya, Maria G. Jafari, Michael Elad, Rémi Gribonval, and Mark D. Plumbley, "Audio Inpainting," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 922–932, Mar. 2012.
- [6] Ondřej Mokřý, Pavel Závíška, Pavel Rajmíc, and Vítězslav Veselý, "Introducing SPAIN (SParse Audio INpainter)," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, IEEE.
- [7] Ondřej Mokřý and Pavel Rajmíc, "Audio inpainting: Revisited and reweighted," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2906–2918, 2020.
- [8] Tomoro Tanaka, Kohei Yatabe, and Yasuhiro Oikawa, "PHAIN: Audio inpainting via phase-aware optimization with instantaneous frequency," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 4471–4485, 2024.
- [9] Ondřej Mokřý, Paul Magron, Thomas Oberlin, and Cédric Févotte, "Algorithms for audio inpainting based on probabilistic nonnegative matrix factorization," *Signal Processing*, p. 10, Dec. 2022.
- [10] Mathieu Lagrange, Sylvain Marchand, and Jean-Bernard Rault, "Long interpolation of audio signals using linear prediction in sinusoidal modeling," *Journal of the Audio Engineering Society*, vol. 53, no. 10, pp. 891–905, 2005.
- [11] Nathanael Perraudin, Nicki Holighaus, Piotr Majdak, and Peter Balazs, "Inpainting of long audio segments with similarity graphs," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018.
- [12] Eloi Moliner and Vesa Välimäki, "Diffusion-based audio inpainting," *Journal of the Audio Engineering Society*, vol. 72, no. 3, pp. 100–113, Mar. 2024.
- [13] Federico Miotello, Mirco Pezzoli, Luca Comanducci, Fabio Antonacci, and Augusto Sarti, "Deep Prior-Based Audio Inpainting Using Multi-Resolution Harmonic Convolutional Neural Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pp. 1–11, 2023.
- [14] Andrés Marafioti, Nathanaël Perraudin, Nicki Holighaus, and Piotr Majdak, "A context encoder for audio inpainting," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2362–2372, Dec. 2019.
- [15] Andres Marafioti, Piotr Majdak, Nicki Holighaus, and Nathanaël Perraudin, "GACELA: A generative adversarial context encoder for long audio inpainting of music," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 1, pp. 120–131, Jan. 2021.
- [16] Alessandro Ilic Mezza, Matteo Amerena, Alberto Bernardini, and Augusto Sarti, "Hybrid packet loss concealment for real-time networked music applications," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 266–273, 2024.
- [17] Georg Tauböck, Shristi Rajbamshi, and Peter Balazs, "Dictionary learning for sparse audio inpainting," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 1, pp. 104–119, Jan. 2021.
- [18] Pavel Závíška, Pavel Rajmíc, and Ondřej Mokřý, "Multiple Hankel matrix rank minimization for audio inpainting," in *2023 46th International Conference on Telecommunications and Signal Processing (TSP)*, jul 2023, IEEE.
- [19] Ismo Kauppinen and Kari Roth, "Audio Signal Extrapolation – Theory and Applications," in *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02)*, Hamburg, Germany, Sept. 2002, pp. 105–110.
- [20] Kari Roth, Ismo Kauppinen, Paulo A. A. Esquef, and Vesa Välimäki, "Frequency warped Burg's method for AR-modeling," in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, 2003, ASPAA-03, IEEE.
- [21] Peter J. Brockwell and Richard A. Davis, *Time Series: Theory and Methods*, Springer series in statistics. Springer, 2nd edition, 2006.
- [22] Udo Zölzer, *DAFX: Digital Audio Effects*, Wiley, 2nd edition, 2011.
- [23] James Durbin, "The fitting of time-series models," *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, vol. 28, no. 3, pp. 233, 1960.
- [24] Norman Levinson, "The Wiener (root mean square) error criterion in filter design and prediction," *Journal of Mathematics and Physics*, vol. 25, no. 1-4, pp. 261–278, Apr. 1946.
- [25] John Parker Burg, *Maximum entropy spectral analysis*, Ph.D. thesis, Stanford University, May 1975.
- [26] John G. Proakis and Dimitris G. Manolakis, *Digital signal processing*, Prentice Hall, 1996.
- [27] Bisrat Derebssa Dufera, Koen Eneman, and Toon van Waterschoot, "Missing sample estimation based on high-order sparse linear prediction for audio signals," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sept. 2018, IEEE.
- [28] Bisrat Derebssa Dufera, Eneyew Adugna, Koen Eneman, and Toon van Waterschoot, "Restoration of click degraded speech and music based on high order sparse linear prediction," in *2019 IEEE AFRICON*, Sept. 2019, IEEE.
- [29] "EBU SQAM CD: Sound quality assessment material recordings for subjective tests," online, 2008.
- [30] Frederic J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978.
- [31] R. Huber and B. Kollmeier, "PEMO-Q—A new method for objective audio quality assessment using a model of auditory perception," *IEEE Trans. Audio Speech Language Proc.*, vol. 14, no. 6, pp. 1902–1911, Nov. 2006.
- [32] Thildo Thiede, William C. Treurniet, Roland Bitto, Christian Schmidmer, Thomas Sporer, John G. Beerends, Catherine Colomes, Michael Keyhl, Gerhard Stoll, Karlheinz Brandenburg, and Bernhard Feiten, "PEAQ – The ITU standard for objective measurement of perceived audio quality," *Journal of the Audio Engineering Society*, vol. 48, no. 1/2, pp. 3–29, January/February 2000.
- [33] P. Kabal, "An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality," Technical Report, MMSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University, May 2002. Available online: <https://www-mmshp.ece.mcgill.ca/Documents/Reports/2002/KabalR2002v2.pdf>
- [34] Ondřej Mokřý, *Modern optimization methods for interpolation of missing sections in audio signals*, Ph.D. thesis, Brno University of Technology, 2024. Available online: <https://www.vut.cz/en/students/final-thesis/detail/160689>
- [35] Juan J. Bosch, Jordi Janer, Ferdinand Fuhrmann, and Perfecto Herrera, "A comparison of sound segregation techniques for predominant instrument recognition in musical audio signals," *International Society for Music Information Retrieval Conference*, pp. 559–564, 2012.
- [36] Juan J. Bosch, Ferdinand Fuhrmann, and Perfecto Herrera, "IRMAS: A dataset for instrument recognition in musical audio signals," 2014. DOI: 10.5281/ZENODO.1290750
- [37] Myles Hollander and Douglas A. Wolfe, *Nonparametric statistical methods*, Wiley series in probability and statistics. Texts and references section. John Wiley & Sons, New York, 2nd edition, 1999.
- [38] "Recommendation ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems," 2015.
- [39] Michael Schoeffler, Sarah Bartoschek, Fabian-Robert Stöter, Marlene Roess, Susanne Westphal, Bernd Edler, and Jürgen Herre, "web-MUSHRA — A comprehensive framework for web-based listening tests," *Journal of Open Research Software*, vol. 6, no. 1, pp. 8, Feb. 2018.