Beta-Divergence-Based Recurrence Plots for Audio Time-Series Analysis

Elena Deckert[1], Philippe Dreesen[1], Norbert Marwan[2], Martijn Boussé[1]
[1]Department of Advanced Computing Sciences (DACS), Maastricht University, Maastricht, The Netherlands
[2]Potsdam Institute for Climate Impact Research, Potsdam, Germany
Email: {m.bousse, philippe.dreesen}@maastrichtuniversity.nl

***Abstract*—While recurrence plots (RPs) are a well-known tool in various fields such as physics, astronomy, and health sciences, their application in audio signal processing remains limited. RPs are a data-analysis tool that visualizes recurrences of states, which are typically measured by the Euclidean norm. When analyzing audio data, however, β-divergences are more common than the Euclidean norm because of their adaptability and suitability for audio-specific characteristics. Therefore, we propose the use of β-divergence-based RPs to gain additional insight into audio data. In this paper, we explore the properties of such RPs, providing a fundamental understanding of their characteristics and an indication of possible future applications. Our findings show that β-divergence-based RPs can provide additional information over traditional RPs, making them well-suited for audio analysis.**

***Index Terms*—recurrence plots, β-divergences, audio analysis**

I. Introduction

Recurrence Plots (RPs) are a versatile visual data-analysis tool for analyzing the dynamical behavior of nonlinear systems with applications in fields ranging from physics and astronomy to health and life sciences [1]–[3]. Recently, there has been an increase in the use of RPs for various applications, driven by the wider availability of relevant software [1].

Different data types have varying characteristics that are not always adequately captured by the Euclidean distance which is typically used in RPs, making it necessary to employ specific distance measures, see [4]. In this paper, we propose the use of β-divergences for the recurrence computation, leading to the definition of β-divergence-based RPs (β-RPs). This flexible method allows us to analyze a wider range of data types, and, in particular, audio data. We show that β-RPs can reveal more information than what is possible with traditional RPs.

Recurrences are repetitions where states are arbitrarily close after a period of time. Visualizing these recurrences with RPs enables the identification of diverse dynamical behaviors and complex patterns within systems [5]. The state space of a univariate time series u of length K can be reconstructed by embedding the time series, using an embedding dimension m and time delay τ, expressed in samples, such that $\mathbf{x}_i = \begin{bmatrix} u_i & u_{i+\tau} & \cdots & u_{i+(m-1)\tau} \end{bmatrix}^{\mathrm{T}}$, where $\mathbf{x}_i$ is the ith embedding of length m using time delay τ and $i \in \{1, 2, \ldots, K-(m-1)\tau\}$. RPs are then constructed by computing the pairwise distance $d(\mathbf{x}_i, \mathbf{x}_j)$ between states. The recurrence matrix $\mathbf{R}$ is defined as

$$r_{ij} = \begin{cases} 1, & \text{if } \ d(\mathbf{x}_i, \mathbf{x}_j) < \epsilon, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } 1 \leq i, j \leq N \tag{1}$$

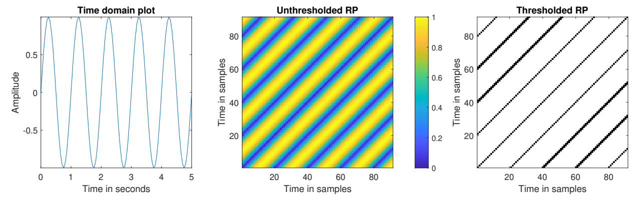


Fig. 1. The parallel diagonal lines in the distance matrix and RP indicate the periodic behavior of the sine function, shown for $f(t) = \sin(2\pi t)$ in [0, 5], sampling $K = 100$ points with $m = 10$, $\tau = 1$, and $\epsilon = 0.1$, comparing $N = 91$ states.

where $\mathbf{x}_i$ is the state of a system at time i, ϵ is the threshold parameter, $d(\cdot, \cdot)$ is the distance metric (usually the Euclidean norm), and $N = K - (m-1)\tau$ is the number of states. The matrix is plotted with a dot at every (i, j)-th entry where $\mathbf{x}_i$ and $\mathbf{x}_j$ are sufficiently close (i.e., a recurrence occurs), revealing informative patterns that relate to properties of the underlying system, see Figure 1 for an example. It can be useful to plot the distance matrix $\mathbf{D}$, where $d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$, using a color scale displaying the full variety in distances.

In addition to the qualitative analysis of RPs, Recurrence Quantification Analysis (RQA) entails the construction of features quantifying diagonal and vertical line structures in RPs [5]. These features can then be used for further analysis or classification tasks. This paper uses the following traditional RQAs to provide a comprehensive framework for characterizing recurrence properties. Recurrence Rate (RR) represents the percentage of recurrence points. Determinism (DET) quantifies the proportion of recurrence points that form diagonal lines at least size ℓ_{min}. The Average Diagonal Length (L) measures the average length of diagonal lines at least size ℓ_{min}. Similarly, Laminarity (LAM) captures the proportion of recurrence points that form vertical lines at least size ℓ_{min}. Finally, Trapping Time (TT) represents the average length of vertical lines at least size ℓ_{min}. RQA calculations depend on the threshold parameter ϵ and the minimum line length ℓ_{min}.

Methods that are similar in spirit to RPs have been applied to music audio, such as self-similarity [6] and KL-based independent subspace analysis [7]. While current applications predominantly use the Euclidean distance to determine recurrences, there have been suggestions to tailor the recurrence definition to the application by using specific similarity metrics [1]. In cases where only phase differences are of interest, e.g., in acoustic signal analysis, the angular distance is recommended [8]. For event-like data, an edit distance metric is used [9]. Uncertain data is addressed by combining RPs with a Bayesian approach to derive probabilities of recurrences [10]. These alternative recurrence definitions suggest the potential adaptability and diversification of the recurrence definition for specific applications. In audio analysis, Itakura-Saito (IS) or Kullback-Leibler (KL) divergence is commonly applied [11]. They are shown to be particularly useful when the data consists of entries of different magnitudes [12]. For example, in nonnegative matrix factorization (NMF) for audio spectra, one can show that more information can be extracted when using the IS- or KL-divergence [11].

The contribution of this paper is RP-based audio analysis using β-divergences, which are a more appropriate metric than conventional Euclidean distance for audio time series, leading to more insights than what is possible with existing methods. In particular, we show that the asymmetry in β-RPs can be a potential source of additional information and leads to new RQA measures. We develop an extensive understanding of characteristics of simple time series and then apply it to audio time series, revealing that the selection of β impacts the emphasis on specific sets of values, highlighting the significance of parameter selection.

II. Methodology

Instead of using "traditional" RPs, which use Euclidean distance to determine recurrences, we compute the pairwise distances in the recurrence matrix using β-divergences. The latter are a continuous interpolation between the IS-divergence ($\beta = 0$), KL-divergence ($\beta = 1$), and least-squares (LS) distance ($\beta = 2$) [12]. They are a special class of Bregman divergences [13] and defined on $\mathbb{R}_+\backslash\{0\}$ for $\beta \in \mathbb{R}$ as

$$d_\beta(x, y) = \begin{cases} \frac{x}{y} - \log(\frac{x}{y}) - 1, & \beta = 0 \\ x(\log x - \log y) + (x - y), & \beta = 1 \\ \frac{x^\beta + (\beta-1)y^\beta - \beta x y^{\beta-1}}{\beta(\beta-1)}, & \beta \in \mathbb{R}\backslash\{0, 1\}. \end{cases}$$

When applied to vectors, the divergence is defined as an element-wise β-divergence on $(\mathbb{R}_+\backslash\{0\})^N$:

$$D_\beta(\mathbf{x}||\mathbf{y}) = \sum_{n=1}^{N} d_\beta(x_n|y_n). \quad (2)$$

Generally, β-divergences with $\beta < 2$ emphasize differences in small values more heavily than $\beta = 2$, while $\beta > 2$ emphasizes differences in large values [12]. Also, KL- or IS-divergence assumes Poisson-distributed data and data with multiplicative Gamma noise, resp., while LS distance assumes additive independent and identically distributed Gaussian noise [11]. The β-divergence for $\beta = 0$ is scale-invariant.

Unlike traditional similarity metrics such as Euclidean distance, β-divergences are *not symmetric*. The LS distance, $\beta = 2$, on the other hand, is symmetric and has the same characteristics as Euclidean distance. LS distance sums the squared differences without taking the square root and includes a factor of $1/2$. Moreover, $\beta = 1$ may compute negative values, but in the application of RPs, we are not interested in the direction of the distance, so the absolute distance is used. β-divergences are limited to distance calculations on positive values. When considering audio data, we typically convert the signal to a positive domain such as the absolute values or square values. All computations were done in MATLAB[1].

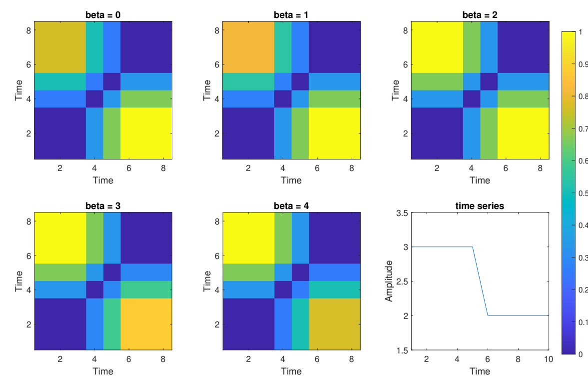


Fig. 2. As we deviate from $\beta = 2$, the β-divergence-based distance matrices exhibit more asymmetric behavior, illustrated for a simple times series [3, 3, 3, 3, 3, 2, 2, 2, 2, 2] using $m = 3$, $\tau = 1$.

III. Experiments and Results

In order to understand the general characteristics of β-RPs, we perform a qualitative analysis on a series of increasingly intricate test signals. We begin with three very simple signals to build intuition about value behavior and allow for easy verification. Complexity is gradually increased, with the third signal introducing a more elaborate structure that begins to resemble audio-like patterns, though the values remain simple. Signals and parameters are chosen in such a way as to demonstrate certain properties and behaviors while keeping them simple enough to maintain comprehensibility.

A. Asymmetry

As β deviates from 2, i.e., as we deviate from the LS distance-based case, asymmetric patterns in the RP become more pronounced. To illustrate this, we analyze a simple signal $u = [3, 3, 3, 3, 3, 2, 2, 2, 2, 2]$ using an embedding dimension of $m = 3$ and a time delay of $\tau = 1$, see Figure 2. When two identical embedding vectors are compared, their distance is 0, which is visible along the diagonal of the plot and, in this case, across most of quadrants 1 and 3. As the embedding vectors differ by 1, 2, or 3 digits, the distances increase accordingly. For $\beta = 2$, the plot is symmetric. However, for $\beta \neq 2$, this symmetry is lost, resulting in a distinct difference between the upper and lower halves of the plot. This is due to the fact that $(x > y) \rightarrow (d_\beta(x|y) > d_\beta(y|x))$ holds for $\beta \in 0, 1$ and $(x > y) \rightarrow (d_\beta(y|x) > d_\beta(x|y))$ holds for $\beta \in 3, 4$. This characteristic is amplified as β deviates from 2.

[1]MATLAB code is publicly available from gitlab.com/mbousse/beta-divergence-based-recurrence-plots-for-audio-time-series-analysis.

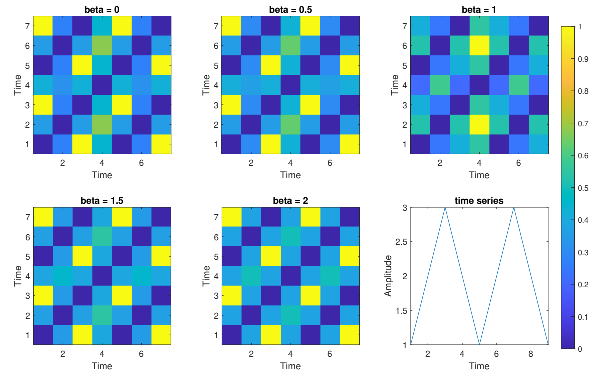


Fig. 3. For $\beta = 1$, the distance matrix deviates as there is less emphasis on the order of values within an embedding, illustrated for a simple time series [1, 2, 3, 2, 1, 2, 3, 2, 1] using $m = 3$, $\tau = 1$.

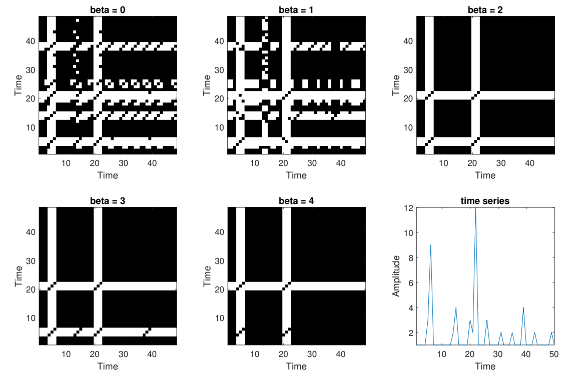


Fig. 4. For $\beta < 2$, the RPs show more emphasis on the differences in small values, illustrated for a time series [1, 1, 1, 1, 3, 9, 1, 1, 1, 1, 1, 1, 1, 1, 2, 4, 1, 1, 1, 1, 3, 2, 12, 1, 1, 1, 3, 1, 1, 1, 1, 2, 1, 1, 1, 2, 1, 1, 1, 4, 1, 1, 1, 2, 1, 1, 1, 1, 1, 2, 1] using $m = 3$, $\tau = 1$, $\epsilon = 0.1$ where $r(i, j) = 0$ is black.

B. Less Emphasis on Value Order for $\beta = 1$

For β-RPs using $\beta = 1$ a deviating RP-pattern can be observed as there is less emphasis on the order of values within an embedding. This can be illustrated for a sawtooth time series (Figure 3). For $\beta \neq 1$ the most distant points appear at coordinates $(1, 7), (1, 3), (3, 1), (3, 5), (5, 3), (5, 7), (7, 1)$, and $(7, 5)$, visualizing the distance of vectors $[1, 2, 3]$ and $[3, 2, 1]$. Looking at $\beta = 1$, these same points have a distance smaller than 1 and are therefore not the most distant points in the plot. The second most distant points for $\beta \neq 1$, at coordinates $(4, 2), (4, 6)$, become the most distant elements for $\beta = 1$. Therefore, in contrast to the other β values, we can observe that the calculation for $\beta = 1$ results in less emphasis on the value order and creates a deviating RP pattern.

C. Emphasis on Differences in Small Values for $\beta < 2$

RP computations with $\beta \in [0, 1]$ visualize the emphasis on differences in small values. In Figure 4, we increased the complexity of the time series by considering multiple peaks. This structure mimics an absolute or square value audio signal that we are analyzing. In the RPs, for $\beta \geq 2$, we can only see the two major peaks, while $\beta = 0$ and $\beta = 1$ show smaller existing peaks. This indicates how KL-divergence and IS-divergence emphasize differences in small values. If we scale this signal with any factor, meaning we apply an amplitude shift, the RPs for all β-values stay the same. If we shift the signal up with a vertical translation, $\beta = 0$ does not see small peaks as small values anymore and therefore does not visualize them. On the contrary, $\beta = 1$ still shows multiple peaks.

D. Scale Invariance

The property of scale invariance that is unique for $\beta = 0$ in β-divergences is not significant in RPs. This is because RPs inherently visualize distances based on the scale of the minimum and maximum distances in the data, ensuring that scale invariance is maintained for all values of β. Consequently, the visual representation of the RP remains unchanged regardless of the value of β. However, for $\beta \neq 0$ the actual distances in an unscaled RP will differ. Therefore we can not uniquely leverage the property of scale invariance for $\beta = 0$.

IV. Stuttering Audio Data Application

We illustrate the use of β-RPs on a stuttering audio dataset, demonstrating that β-RPs can exhibit more information than traditional RPs. The audio signals are taken from the SEP-28k-E stutter podcast dataset [14], [15]. Recall that β-divergences can only be used on positive values, hence, we first square the values. The signals are also downsampled to reduce computational load. A common lower limit on the sampling rate for human speech audio is 8 kHz, which is sufficient to capture speech frequencies up to 4 kHz [16]. We decided to downsample the data to a rate of 2 kHz as this allows the speech signals to still be comprehensible when listening to it.

In Figure 5, the audio signal has two sets of large peaks and multiple ones with medium peaks. The β-divergence-based distance matrices for $\beta \geq 2$ strongly emphasize the larger sets of peaks while $\beta = 1$ also visualizes the smaller ones, demonstrating suitability for audio signals with great range in amplitude. For the second signal, in Figure 6, this characteristic is not quite as relevant, as the signal mostly has sets of high peaks. We can also see that $\beta = 1$ has fewer vertical lines due to its asymmetric nature.

A. Importance of Parameter Selection for Audio Data

Parameter selection in RP computations is crucial to capture relevant characteristics of audio signals. The analysis strongly depends on the embedding dimension m, time delay τ, and threshold ϵ. Increasing m expands the vertical and horizontal line structures, representing distant parts in the time series (Figure 7). This behavior also appears in β-RPs for all values of β. A larger time delay increases the non-recurrent streaks, especially in the lower right half, see Figure 8.

A suitable threshold ϵ is crucial to reveal relevant features in the RP as shown in Figure 9. With a higher value, e.g. 0.1, we recognize only the large peaks in the time series. In contrast,

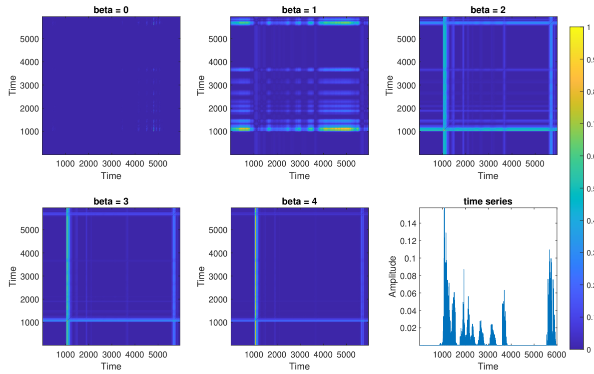


Fig. 5. For $\beta = 1$ the distance matrix recognizes smaller peaks in the time series, illustrated for an audio signal StutterTalk_2_15 with $m = 50$, $\tau = 1$.

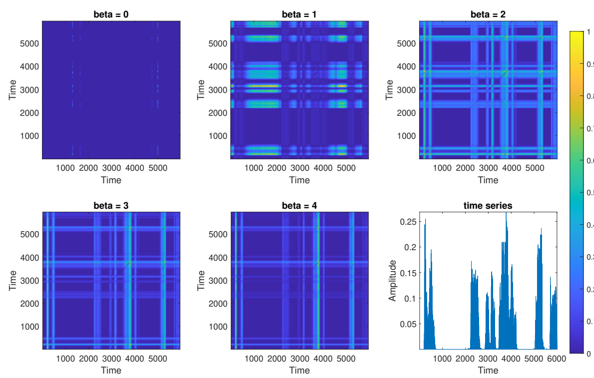
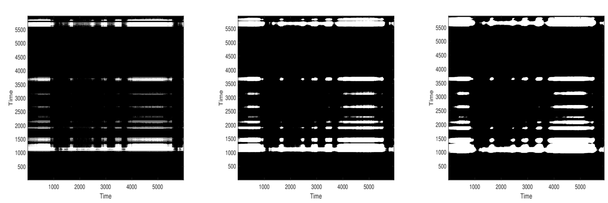
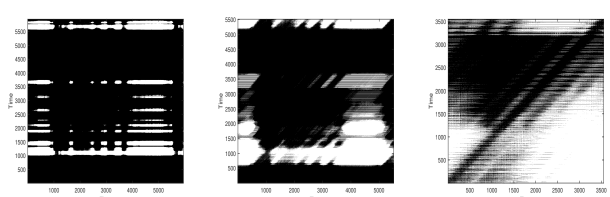


Fig. 6. For $\beta = 1$ the distance matrix shows fewer vertical streaks, illustrated for an audio signal StutterTalk_2_19 with $m = 50$, $\tau = 1$.



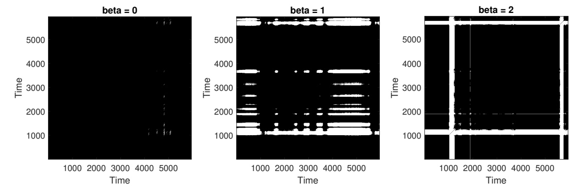
(a) $m = 10$ (b) $m = 50$ (c) $m = 100$

Fig. 7. Increasing the embedding dimension increases the width of horizontal distant streaks in the RP, illustrated for an audio signal StutterTalk_2_15 with $\tau = 1$, $\epsilon = 0.1$, $\beta = 1$, where $r(i, j) = 0$ is black.

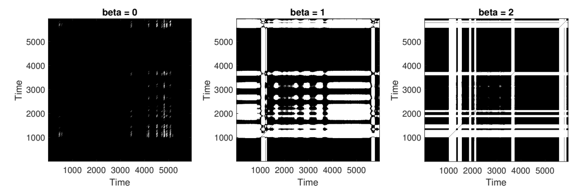


(a) $\tau = 1$ (b) $\tau = 10$ (c) $\tau = 50$

Fig. 8. Increasing the time delay increases the number of non-recurrent embeddings in the RP, illustrated for an audio signal StutterTalk_2_15 with $m = 50$, $\epsilon = 0.1$, $\beta = 1$, where $r(i, j) = 0$ is black.



(a) Threshold $\epsilon = 0.1$



(b) Threshold $\epsilon = 0.02$

Fig. 9. Threshold value has a strong influence on the information we can retrieve from the RP, illustrated for an audio signal StutterTalk_2_15 with $m = 50$, $\tau = 1$, where $r(i, j) = 0$ is black.

TABLE I
RQA VALUES FOR DIFFERENT β VALUES CALCULATED FOR AN AUDIO SIGNAL STUTTERTALK_2_15.WAV, WITH $\epsilon = 0.1$

β	RR	DET	L	LAM	LAM_h	ΔLAM	TT	TT_h	ΔTT
0	0.9992	0.9997	2878.33	0.9973	0.9961	0.0012	3639.98	5034.93	1394.95
1	0.8732	0.9792	830.04	0.9875	0.9751	0.0124	2711.13	969.40	1742.73
2	0.8744	0.9924	785.44	0.9935	0.9935	0	1395.80	1395.80	0
3	0.9210	0.9904	998.14	0.9987	0.9924	0.0063	1860.58	1897.69	37.11
4	0.9451	0.9878	1194.76	0.9951	0.9921	0.0030	1884.90	2891.45	1006.55

a smaller value, e.g. 0.02, reveals more spikes, potentially exposing relevant information. For $\beta = 1$, decreasing the threshold also reinstates the vertical pattern.

B. RQA in Audio Data β-RPs

The emphasis on small values and the loss of vertical non-recurrent streaks is also clear from the RQA. Moreover, we introduce possible new *horizontal* RQA measures. Table I shows the RQA for the RPs in Figure 5 with $\ell_{\text{min}} = 100$ and $\epsilon = 0.1$. We added horizontal features where "LAM_h" represents the proportion of recurrent points that form horizontal lines at least size ℓ_{min} and "TT_h" the average horizontal line length. For the symmetric RP, $\beta = 2$, these are equivalent to LAM and TT. The RR for $\beta = 0$ is very high, followed by $\beta = 3, 4$ and then $\beta = 1, 2$ (Table I). The latter has a similar RR because while $\beta = 1$ has more non-recurrent horizontal lines, $\beta = 2$ has additional vertical lines, see Figure 9. The fluctuation in RR indicates that a threshold selection based on a fixed RR might be useful when using different β values in classification tasks. Moreover, the average line lengths horizontally are shorter for $\beta = 1$ than the vertical ones as we do not have the repetition of vertical non-recurrent streaks. ΔLAM and ΔTT demonstrate the absolute difference in the vertical and horizontal values. These features can be interpreted as an asymmetry measure.

RQA features can be used to train machine learning algorithms for automated stuttering classification. Indeed, RPs could possibly characterize repetitions, which can support speech pathologists with stuttering analysis. Our preliminary results indicate that stuttering classification using RQAs computed from β-RPs can perform better than traditional RPs.

V. Discussion

This work provides an initial indication of the behavior of β-RPs by analyzing simple time series. Notably, when $\beta \neq 2$ we can observe asymmetric behavior that decreases as β approaches 2, where symmetry is restored. This asymmetry reveals different information looking at horizontal and vertical features. Establishing an understanding of these new horizontal features and their connection to the physical process in the underlying system remains intricate and requires deeper investigation. Moreover, the asymmetry of β-RPs as β approaches 0 tends to result in points being closer together. Consequently, a lower threshold might be necessary to obtain valuable insights, meaning that the range for suitable thresholds deviates from existing research. A certain degree of robustness towards the order of values in an embedding dimension was demonstrated for $\beta = 1$, suggesting it could be a suitable measure for counting data. Furthermore, the property of $\beta < 2$ emphasizing differences in small values could be observed.

When using β-RPs for (speech) audio, selecting suitable parameters—threshold, embedding dimension, time delay, and β—is essential to tailor the analysis to specific data properties. Visual inspection shows strong sensitivity to these choices, which can be addressed using selection techniques like False Nearest Neighbors (FNN) for the embedding dimension and Mutual Information for the time delay. Audio can vary greatly in amplitude due to factors such as differences between speakers, variations in pitch, and other acoustic characteristics [17]. Therefore, $\beta = 1$ shows promising results as it also considers the distances in values in a smaller range of amplitudes. In typical RP analysis, thresholds are typically around 0.1. For β-RPs in audio data analysis, optimal thresholds vary for each β-value and are generally lower. Therefore, selecting a threshold that ensures a pre-selected RR could be useful.

VI. Conclusion

In this paper, we demonstrated that β-divergences can be a viable similarity metric for RP construction by showing their potential for audio time series analysis. The asymmetric nature of β-RPs in contrast to traditional RPs can be a source of additional information and leads to the definition of additional RQA measures based on horizontal features. We recommend comparing β-RPs with other (non)-Euclidean distance measures in future work. Furthermore, beyond the normal parameter tuning, the additional numerical value of β can also be used as a hyperparameter to tailor the analysis to specific data characteristics, especially for machine learning algorithms. In future work, we will analyze the potential of β-RPs in audio classification tasks.

References

[1] N. Marwan and K. H. Kraemer, "Trends in recurrence analysis of dynamical systems," *The European Physical Journal Special Topics*, vol. 232, 2023.

[2] Y. Shinchi, N. Takeda, H. Gotoda, T. Shoji, and S. Yoshida, "Early detection of thermoacoustic combustion oscillations in staged multisector combustor," *AIAA Journal*, vol. 59, no. 10, pp. 4086–4093, 2021.

[3] M. Frasch, C. Herry, Y. Niu, and D. Giussani, "First evidence that intrinsic fetal heart rate variability exists and is affected by hypoxic pregnancy," *The Journal of Physiology*, vol. 598, 2020.

[4] N. Marwan, "Challenges and perspectives in recurrence analyses of event time series," *Frontiers in Applied Mathematics and Statistics*, vol. 9, 2023.

[5] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Phys. Rep.*, vol. 438, no. 5, pp. 237–329, 2007.

[6] J. Foote, "Visualizing music and audio using self-similarity," ACM Multimedia, 1999, pp. 77–80.

[7] M. A. Casey and A. Westner, "Separation of mixed audio sources by independent subspace analysis," in *International Computer Music Conference*, 2000.

[8] C. Ioana, A. Digulescu, A. Serbanescu, I. Candel, and F. Birleanu, "Recent advances in non-stationary signal processing based on the concept of recurrence plot analysis," *Springer Proceedings in Mathematics and Statistics*, vol. 103, pp. 75–93, 2014.

[9] S. Suzuki, Y. Hirata, and K. Aihara, "Definition of distance for marked point process data and its application to recurrence plot-based analysis of exchange tick data of foreign currencies," *International Journal of Bifurcation and Chaos*, vol. 20, 2010.

[10] B. Goswami, N. Boers, A. Rheinwalt, *et al.*, "Abrupt transitions in time series with uncertainties," *Nature Communications*, vol. 9, 2018.

[11] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural computation*, vol. 21, pp. 793–830, 2008.

[12] M. Vandecappelle, N. Vervliet, and L. De Lathauwer, "Rank-one tensor approximation with beta-divergence cost functions," in *27th EUSIPCO*, 2019, pp. 1–5.

[13] R. Hennequin, B. David, and R. Badeau, "Beta-divergence as a subclass of bregman divergence," *IEEE Signal Processing Letters*, vol. 18, pp. 83–86, 2011.

[14] S. P. Bayerl, D. Wagner, E. Nöth, T. Bocklet, and K. Riedhammer, "The influence of dataset partitioning on dysfluency detection systems," in *TSD*, 2022.

[15] C. Lea, V. Mitra, A. Joshi, S. Kajarekar, and J. P. Bigham, *SEP-28k: A dataset for stuttering event detection from podcasts with people who stutter*, ICASSP 2021.

[16] D. Jurafsky and J. H. Martin, *Speech and Language Processing*. Prentice Hall, 2000.

[17] Y. Lee, P. Keating, and J. Kreiman, "Acoustic voice variation within and between speakers," *J. Acoust. Soc. Am.*, vol. 146, pp. 1568–1579, 2019.