Fraunhofer FKIE

FRAUNHOFER INSTITUTE FOR COMMUNICATION, INFORMATION PROCESSING AND ERGONOMICS FKIE

Robust Detection of Jittered Multiply Repeating Audio Events Using Iterated Time-Warped ACF

Frank Kurth, Kevin Wilkinghoff

1. Introduction

Goal

- *Robust* detection of multiply "almost repeating" (jittered) events in time series
- Model: Repeating onset times $(t_0, t_0 + \lambda, \dots, t_0 + K\lambda) + (\delta_0, \delta_1, \dots, \delta_K)$, $\max(|\delta_i|) < \lambda/2$

Approach

- Combine *shift-ACF,* a variant of classical ACF, with *Dynamic Time Warping (DTW)*

Evaluation

- Newly introduced *ITW-ACF* outperforms ACF and shift-ACF on bioacoustics datasets

2. Shift-Method and Shift-ACF

Shift-product: (for integer s)

$\mathbb{O}_s^0[x](k) \coloneqq P_s[x](k) \coloneqq x(k) \cdot \overline{x(k-s)}$

emphasizes repeating components.

Binary version: $\mathbb{O}_s^0[x,y](k,l) \coloneqq x(k) \cdot \overline{y(l-s)}$

Shift-minimum:

$\mathbb{O}_s^1[x](k) \coloneqq M_s[x](k) \coloneqq \min(|x(k)|, |x(k-s)|)$

suppresses non-repeating components

General shift-method framework:

$\mathbb{O}_s^t \coloneqq \mathbb{O}_s^{t_1} \circ \cdots \circ \mathbb{O}_s^{t_n}, \quad t = (t_1, \cdots, t_n) \in \{0,1\}^n$

Shift Operations ("ShOps"):

$\mathbb{O}_s^t[x]$ is called ShOp, $(\mathbb{O}_s^t[x])_s$ ShOp-matrix

Shift-ACF of type t:

$$ACF^t[x](s) \coloneqq \sum_{k \in \mathbb{Z}} \mathbb{O}_s^t[x](k)$$

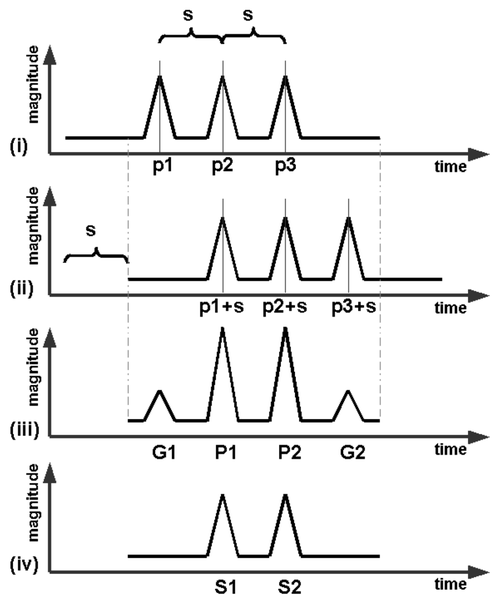


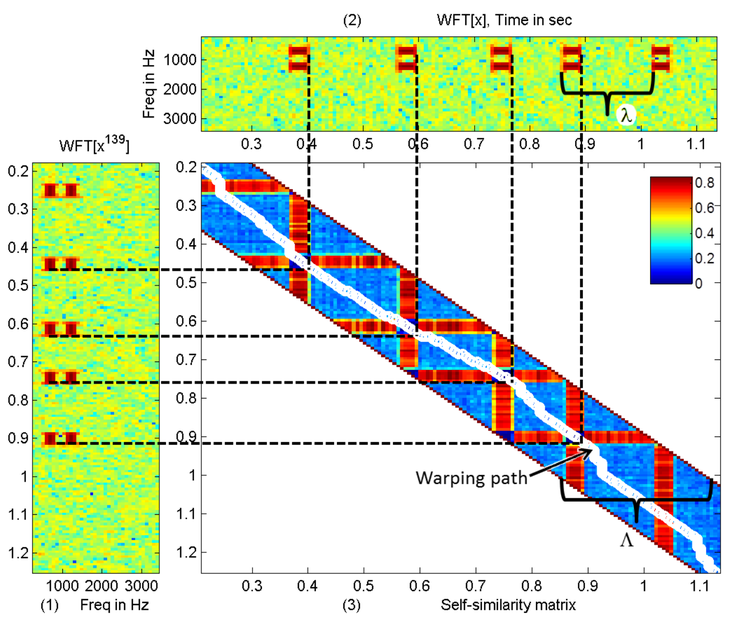
Illustration of shift-operators

3. Jittered Multiply Repeating Events and DTW

(1) Spectrogram $\mathrm{WFT}[x]$ of DTMF Signal x.
x contains multiply "almost repeating" (jittered) events.

(2) Spectrogram of same signal as in (1) delayed by 139 samples, i.e. $\mathrm{WFT}[x^{139}]$.

(3) Warping path between $\mathrm{WFT}[x]$ and $\mathrm{WFT}[x^{139}]$ obtained by DTW restricted to Sakoe-Chiba-band.



4. Iterated Time-Warped ACF (ITW-ACF)

Main Idea

- Perform shift-operations $\mathbb{O}_s^0$ and $\mathbb{O}_s^1$ along warping path between x and x^s

Iteration for fixed shift s along shift type $t = (t_1, \cdots, t_n) \in \{0,1\}^n$

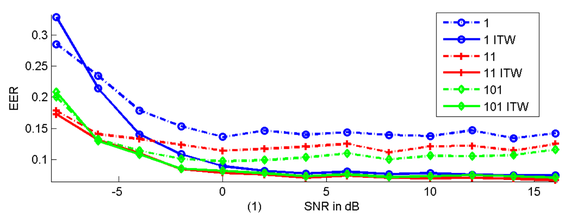
In step $1 \le i \le n$

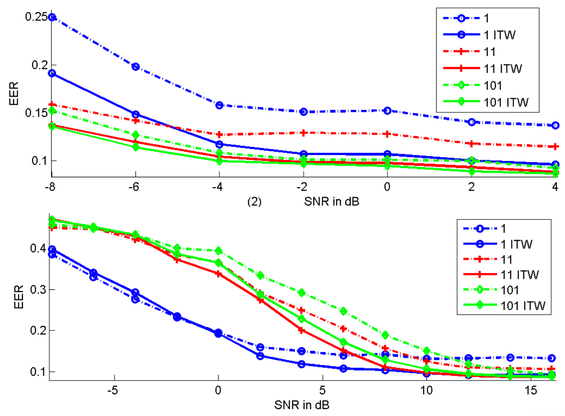
1. Compute band-restricted DTW of x and x^s yielding warping path $((a_1, b_1), \dots, (a_P, b_P))$
2. Compute warped t_i-operation $y(k) \coloneqq \mathbb{O}_0^{t_i}[x, x^s](a_k, b_k)$ for $1 \le k \le P$
3. Unwarp y using back-projection to $(a_1, \dots, a_P)$, yielding warped ShOp $\widetilde{\mathbb{O}}_s^{t_i}[x]$

Iterated time-warped ACF is defined as

$$ITW - ACF^t[x](s) \coloneqq \sum_{k \in \mathbb{Z}} \widetilde{\mathbb{O}}_s^{t_i}[x](k)$$

5. Evaluation





Detection equal error rates (EER) for:

(1) DTMF-signals depending on SNR-level of added Gaussian noise.
(5 DTMF-tones repeated at initial (random) IOI in [80, 200] ms. Jitter: maximum individual deviation of 20 ms.)

(2) Bioacoustic events depending on SNR-level of added Gaussian noise.

(3) bioacoustic events depending on SNR-level of added realistic bioacoustic background noise.

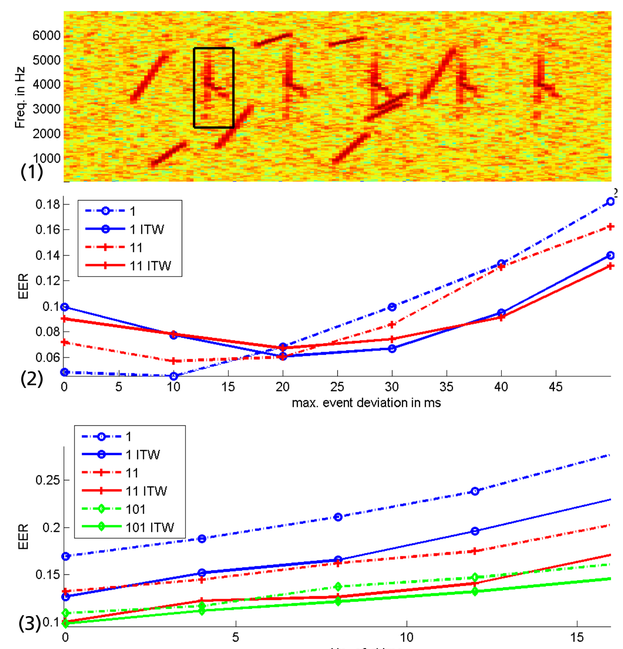
(Counting of TPs and FPs is detailed below, see Evaluation.)

(1) 5-fold repeated bioacoustic event* (framed) with 9 added chirps at 0 dB Gaussian background noise.

(2) Detection-EERs for bioacoustic events with added realistic bioacoustic background noise at 10 dB, depending on maximum event deviation in ms.

(3) Detection-EERs for bioacoustic events with added Gaussian noise at -5 dB for different numbers of added chirp signals.

*Bioacoustic events taken from
- Reference System of Animal Vocalisations of Museum für Naturkunde Berlin, http://www.animalsoundarchive.org/RefSys/
- xeno-canto, https://www.xeno-canto.org/



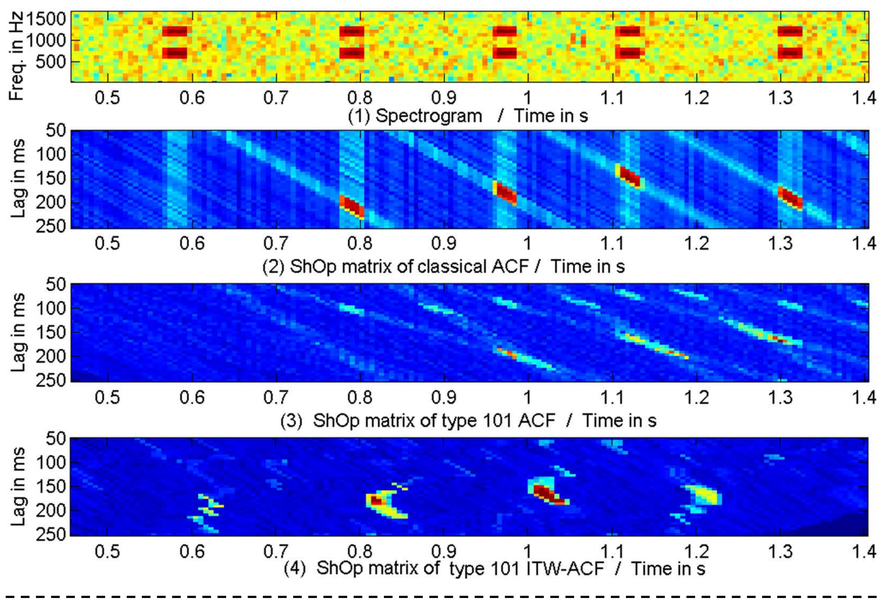
Comparison of standard ACF, Shift-ACF, and proposed Iterated Time-Warped ACF

(1) Spectrogram of signal x containing 5 subsequent DTMF-tones.

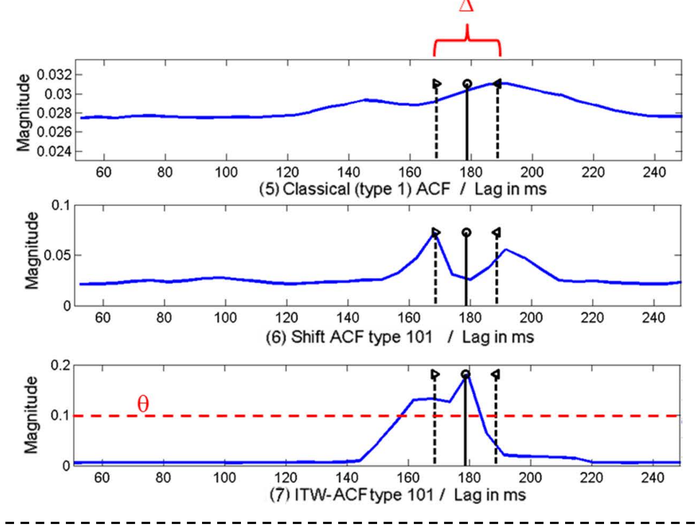
(2) ShOp-matrix $(\mathbb{O}_s^0[x])_s$, i.e., for classical ACF.

(3) ShOp-matrix $(\mathbb{O}_s^{101}[x])_s$.

(4) ShOp-matrix for proposed ITW-ACF, i.e., $(\widetilde{\mathbb{O}}_s^{101}[x])_s$.



Evaluation: TPs are lags inside dashed region of width Δ around known basic IOI (circles) with ACF $> \theta$.
FPs: lags outside dashed region with ACF $> \theta$.



Different ACFs corresponding to (2)-(4):

(5) standard (type 1) ACF

(6) type 101 shift-ACF,

(7) Proposed warped ACF of type 101

Ground truth IOI (circle) and 20 ms tolerance region (dashed lines) are indicated in black.

Note that only ITW-ACF exceeds threshold $\theta = 0.1$

CONTACT: **Frank Kurth** | +49 228 9435-868 | frank.kurth@fkie.fraunhofer.de