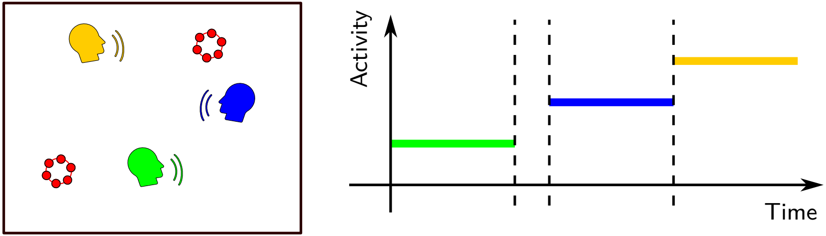
On Synchronization of Wireless Acoustic Sensor Networks in the Presence of Time-varying Sampling Rate Offsets and Speaker Changes

Tobias Gburrek, Joerg Schmalenstroeer, Reinhold Haeb-Umbach

Department of Communications Engineering – Paderborn University

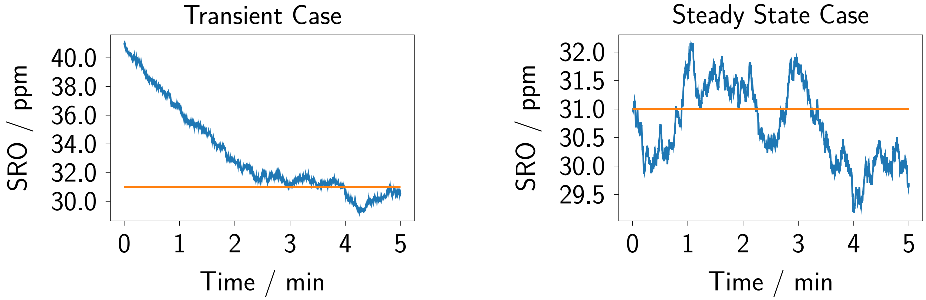
icassp 2022 Singapore

Scenario



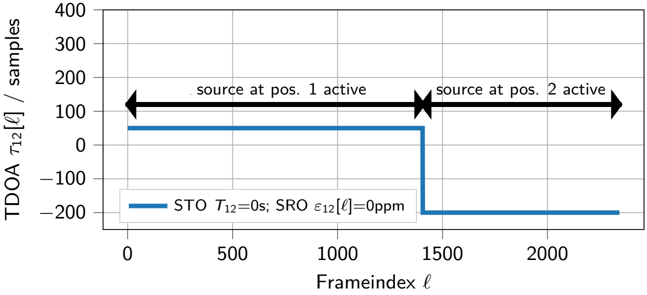
- Independent analog-digital converters (ADCs)
  - Varying frequencies of the oscillators driving the ADCs ⇒ Sampling rate offset (SRO)
  - Varying recording start ⇒ Sampling time offset (STO)
- Previous works: Constant SRO and single fixed source positions
- Here: **Time-varying SRO** and **multiple source positions**

Time-varying SRO



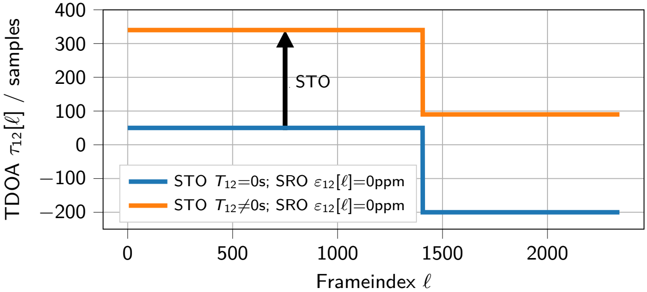
- Transient case: Temperature changes, ...
- Steady state fluctuations: Supply voltage changes, ...

Effect of SRO and STO



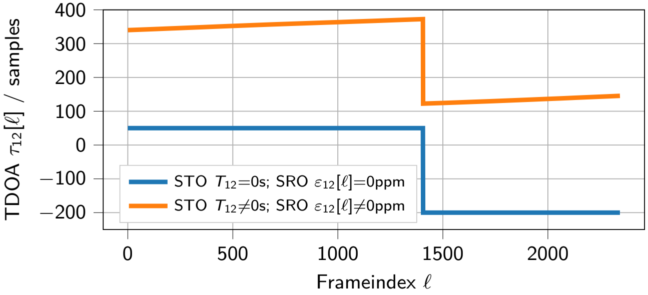
- Perfect synchronization (STO T_{12}=0s; SRO $\varepsilon_{12}[\ell]$=0ppm): Time difference of arrival (TDOA) corresponds to time difference of flight (TDOF)
- STO $\rightarrow$ Constant offset
- Effect of SRO accumulates over time $\rightarrow$ "Ramps"

Effect of SRO and STO



- Perfect synchronization (STO $T_{12}=0$s; SRO $\varepsilon_{12}[\ell]=0$ppm): Time difference of arrival (TDOA) corresponds to time difference of flight (TDOF)
- STO $\rightarrow$ Constant offset
- Effect of SRO accumulates over time $\rightarrow$ "Ramps"

Effect of SRO and STO



- Perfect synchronization (STO $T_{12}=0$s; SRO $\varepsilon_{12}[\ell]=0$ppm): Time difference of arrival (TDOA) corresponds to time difference of flight (TDOF)
- STO $\rightarrow$ Constant offset
- Effect of SRO accumulates over time $\rightarrow$ "Ramps"

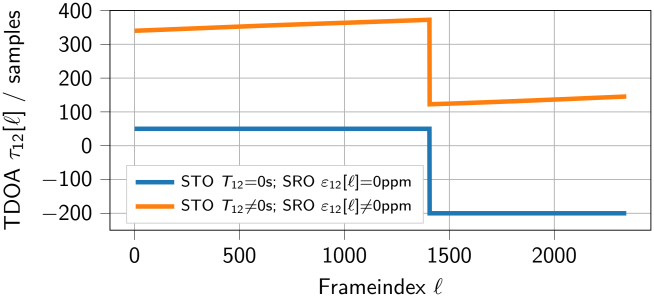
Dynamic Weighted Average Coherence Drift (DWACD)

- Build upon the online weighted average coherence drift (WACD) method [1]
  - Product of consecutive coherence functions: $P_\Gamma(\ell, k) = \Gamma_{12}(\ell, k) \cdot \Gamma^*_{12}(\ell - \ell_d, k)$
  - Under the following conditions:
    - Source position is the same in ℓ-th frame and $(\ell - \ell_d)$-th frame
    - SRO is close to zero

  it holds [1]: $\angle P_\Gamma(\ell, k) \propto \varepsilon_{12}[\ell]$
- Handle time-varying SROs: $\bar{P}_\Gamma(\ell, k) = \alpha \cdot \bar{P}_\Gamma(\ell - 1, k) + (1 - \alpha) \cdot P_\Gamma(\ell, k)$
- Handle source position changes: Reduce temporal context used to calculate the coherence product $P_\Gamma(\ell, k)$
- Generalized cross correlation based SRO estimation from avg. coherence product $\bar{P}_\Gamma(\ell, k)$
- Use current SRO estimate to resample the next signal frame for coherence estimation

[1] A. Chinaev et al., “Online Estimation of Sampling Rate Offsets in Wireless Acoustic Sensor Networks with Packet Loss,”

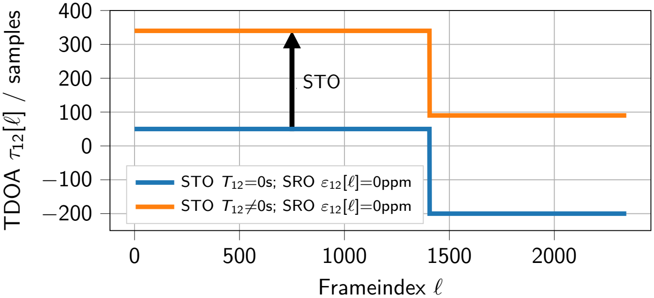
STO Estimation



- Goal: Compensate for STO such that synchronized signals correctly represent the time of flight (TOF) from source position to microphone position
- Principle: Align TDOA with TDOF
  - TDOA after compensating for SRO: $\tau_{12}[\ell] = \underbrace{(d_2[\ell] - d_1[\ell])/c \cdot f_s}_{\text{TDOF}} - \underbrace{(T_2 - T_1) \cdot f_s}_{\tau_{12}^{\text{STO}}}$
  - Estimated source-microphone distances [2] → TDOF estimates

[2] T. Gburrek et al., "On source-microphone distance estimation using convolutional recurrent neural networks"

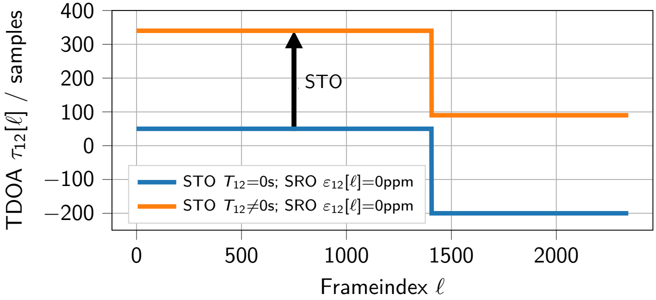
STO Estimation



- Goal: Compensate for STO such that synchronized signals correctly represent the time of flight (TOF) from source position to microphone position
- Principle: Align TDOA with TDOF
  - TDOA after compensating for SRO: $\tau_{12}[\ell] = \underbrace{(d_2[\ell] - d_1[\ell])/c \cdot f_s}_{\text{TDOF}} - \underbrace{(T_2 - T_1) \cdot f_s}_{\tau_{12}^{\text{STO}}}$
  - Estimated source-microphone distances [2] → TDOF estimates

[2] T. Gburrek et al., "On source-microphone distance estimation using convolutional recurrent neural networks"

STO Estimation



- Goal: Compensate for STO such that synchronized signals correctly represent the time of flight (TOF) from source position to microphone position
- Principle: Align TDOA with TDOF
  - TDOA after compensating for SRO: $\tau_{12}[\ell] = \underbrace{(d_2[\ell] - d_1[\ell])/c \cdot f_s}_{\text{TDOF}} - \underbrace{(T_2 - T_1) \cdot f_s}_{\tau_{12}^{\text{STO}}}$
  - Estimated source-microphone distances [2] → TDOF estimates

[2] T. Gburrek et al., "On source-microphone distance estimation using convolutional recurrent neural networks"

Experiments - Data Set

- 100 simulated acoustic sensor networks with 4 nodes
- Pair-wise SRO and STO estimation (first node as reference)
- 5 min long recordings
  - Image-source method for room impulse response simulation
  - Random number of positions M
  - At each position: up to 4 utterances from the TIMIT data base as source signal
  - Speech pauses with a length between 0.5 s and 2 s
  - Sensor noise: average SNR value of 30 dB at a source-node distance of 3.2 m
- Simulation of asynchronous sampling:
  - STO between -1 s and 1 s
  - Steady state SRO between -100 ppm and 100 ppm
  - Steady state standard deviation of the SRO of 1.25 ppm
  - STFT-resampling method for SRO simulation

Comparison of SRO Estimators

	SRO	Position	Method	avg. RMSE(ε) in ppm	avg. RMSE(τ^{SRO}) in samples
Setup-1	Constant	Fixed	Online WACD	0.21	**0.14**
			DXCP-PhaT [3]	**0.15**	0.36
			DWACD	0.40	0.15
Setup-2	Varying	Fixed	Online WACD	0.63	0.73
			DXCP-PhaT	0.66	0.97
			DWACD	0.51	1.10
Setup-3	Varying	Changing	Online WACD	2.80	3.25
			DXCP-PhaT	22.42	16.61
			DXCP-PhaT$_8$	1.28	2.81
			DWACD	0.64	0.32

[3] A. Chinaev et al.,"Double-Cross-Correlation Processing for Blind Sampling-Rate and Time-Offset Estimation"

Comparison of SRO Estimators

	SRO	Position	Method	avg. RMSE(ε) in ppm	avg. RMSE(τ^{SRO}) in samples
Setup-1	Constant	Fixed	Online WACD	0.21	0.14
			DXCP-PhaT [3]	0.15	0.36
			DWACD	0.40	0.15
Setup-2	Varying	Fixed	Online WACD	0.63	0.73
			DXCP-PhaT	0.66	0.97
			DWACD	**0.51**	**0.32**
Setup-3	Varying	Changing	Online WACD	2.80	3.25
			DXCP-PhaT	22.42	16.61
			DXCP-PhaT_8	1.28	2.81
			DWACD	0.64	0.32

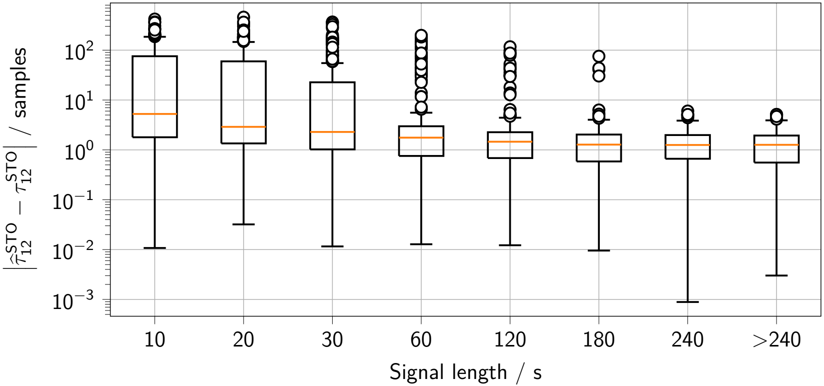
[3] A. Chinaev et al.,"Double-Cross-Correlation Processing for Blind Sampling-Rate and Time-Offset Estimation"

Comparison of SRO Estimators

	SRO	Position	Method	avg. RMSE(ε) in ppm	avg. RMSE(τ^{SRO}) in samples
Setup-1	Constant	Fixed	Online WACD	0.21	0.14
			DXCP-PhaT [3]	0.15	0.36
			DWACD	0.40	0.15
Setup-2	Varying	Fixed	Online WACD	0.63	0.73
			DXCP-PhaT	0.66	0.97
			DWACD	0.51	0.32
Setup-3	Varying	Changing	Online WACD	2.80	3.25
			DXCP-PhaT	22.42	16.61
			DXCP-PhaT$_8$	1.28	2.81
			DWACD	**0.64**	**0.32**

[3] A. Chinaev et al.,“Double-Cross-Correlation Processing for Blind Sampling-Rate and Time-Offset Estimation”

STO Estimation Performance



Summary

- Scenario: Time-varying SROs and source position changes
- SRO estimation via DWACD method
  - Adaptation of our online WACD method to dynamic scenario
- STO estimation method
  - Synchronization which reflects the physical TOFs
  - Usage of source-microphone distance estimates as support information

- **Open source toolbox**: `https://github.com/fgnt/paderwasn`
- Questions?: gburrek@nt.uni-paderborn.de
- Poster: Friday, 13 May, 20:00 - 20:45 (AUD-31: Multichannel Processing)