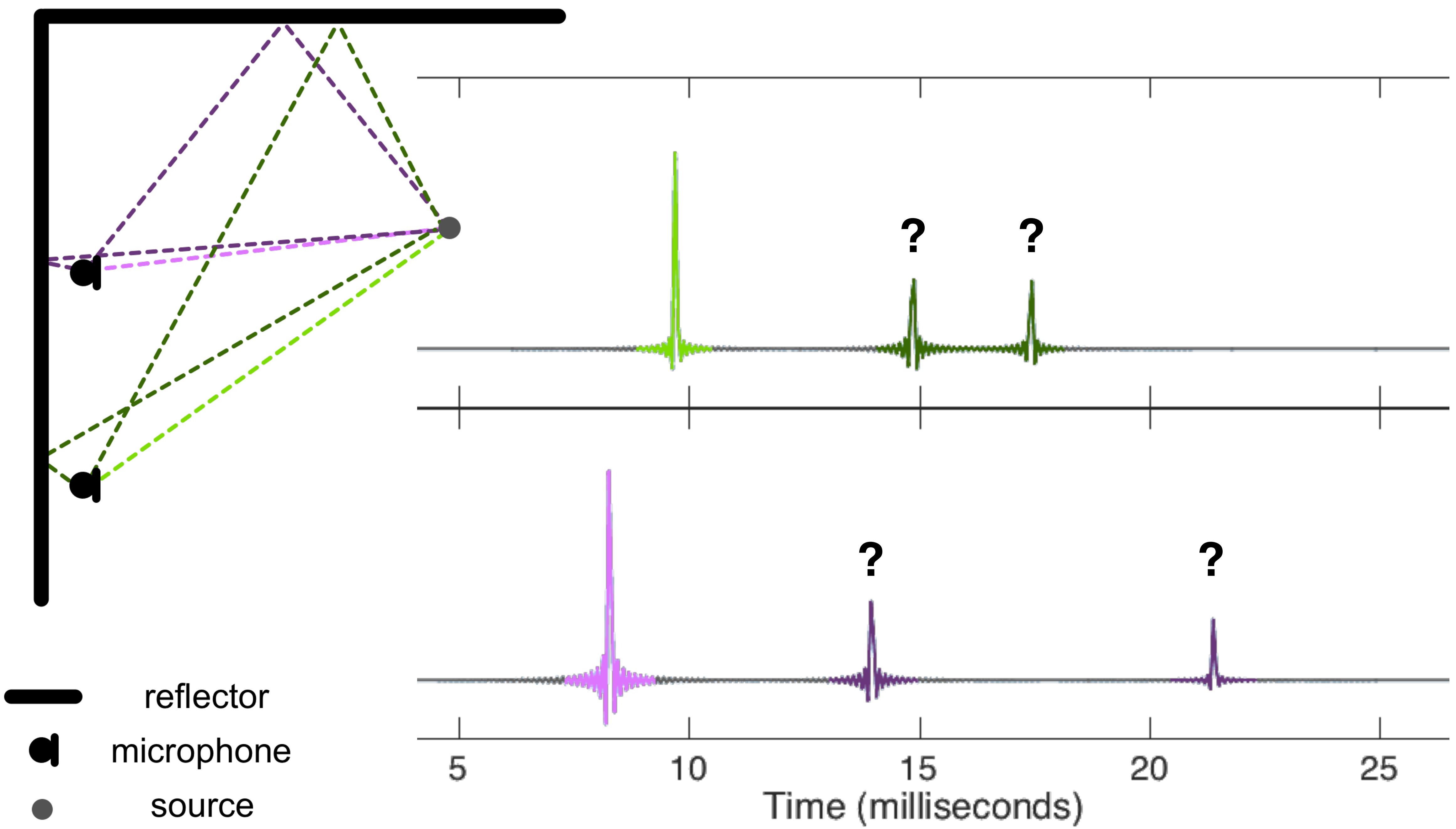
TIME OF ARRIVAL DISAMBIGUATION USING THE LINEAR RADON TRANSFORM

Youssef El Baba, Andreas Walther, Emanuël A.P. Habets

International Audio Laboratories Erlangen
youssef.elbaba@audiolabs-erlangen.de

Challenge


?
?
?
?
reflector
microphone
source
5
10
15
20
25
Time (milliseconds)

Content

Introduction

- Research context: Room Geometry Inference
- Problem: echo labeling
- Focus on (uniform) linear arrays
- Applications:
  - Room geometry inference
  - Blind source separation
  - Sound source localization

Transducer Setup

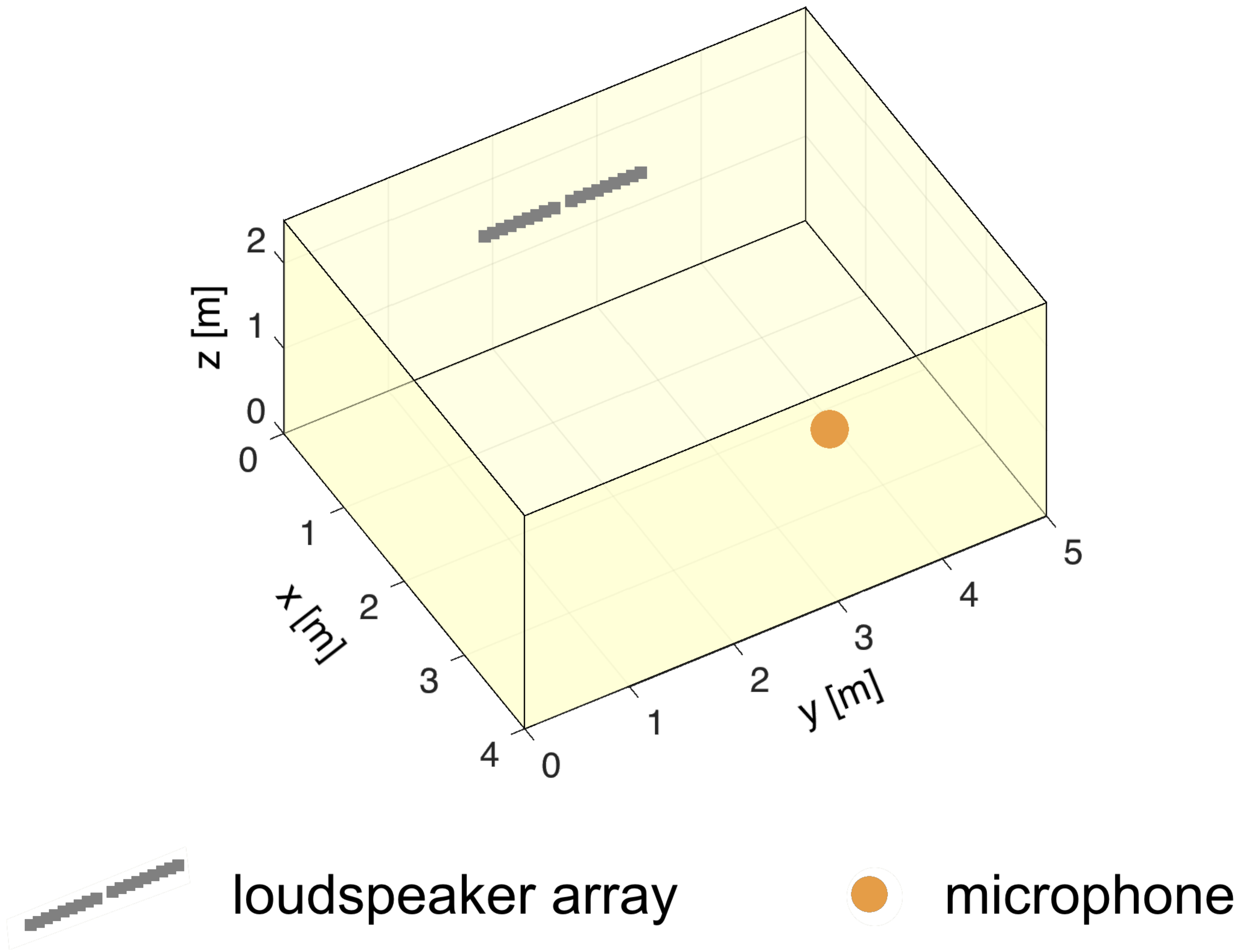
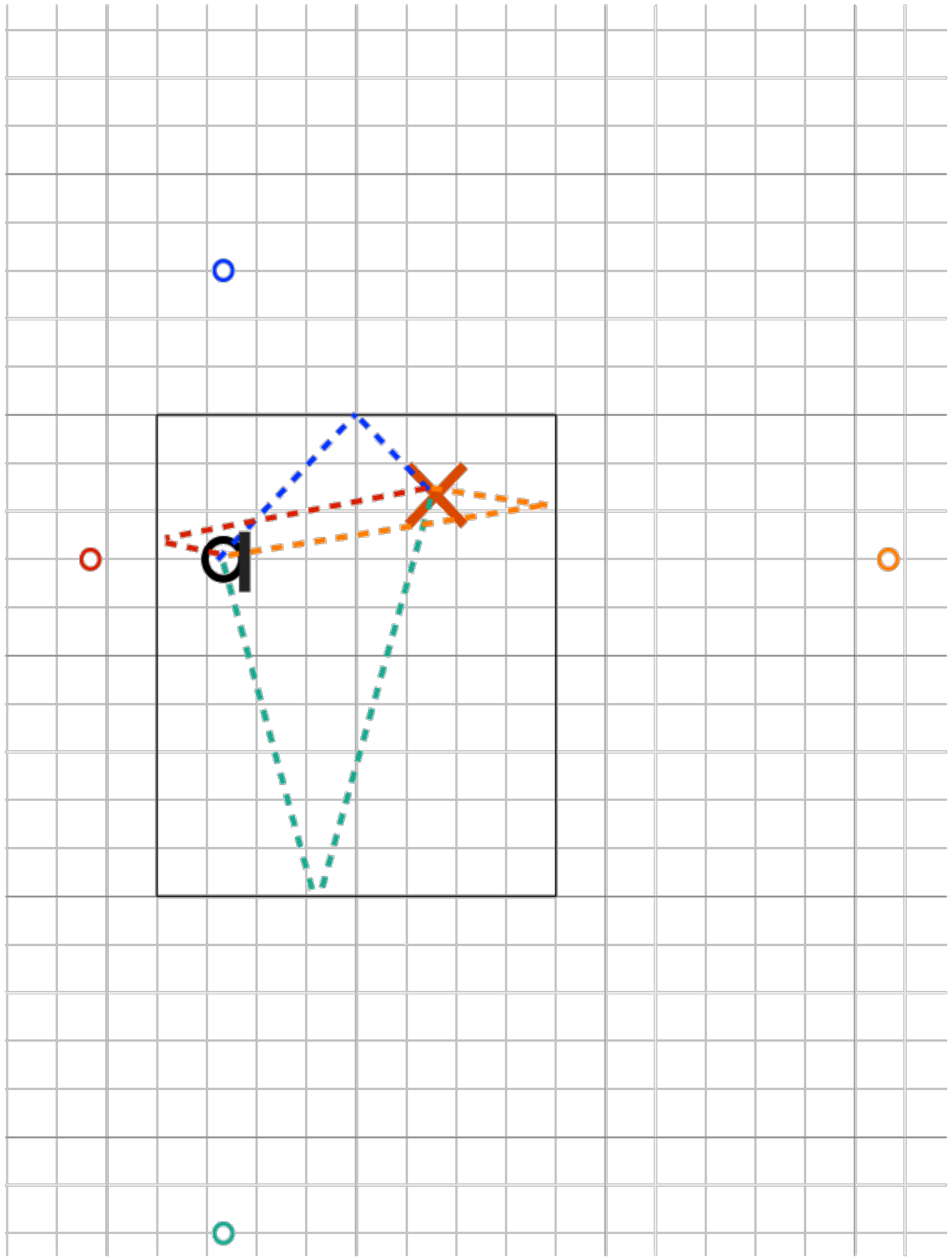
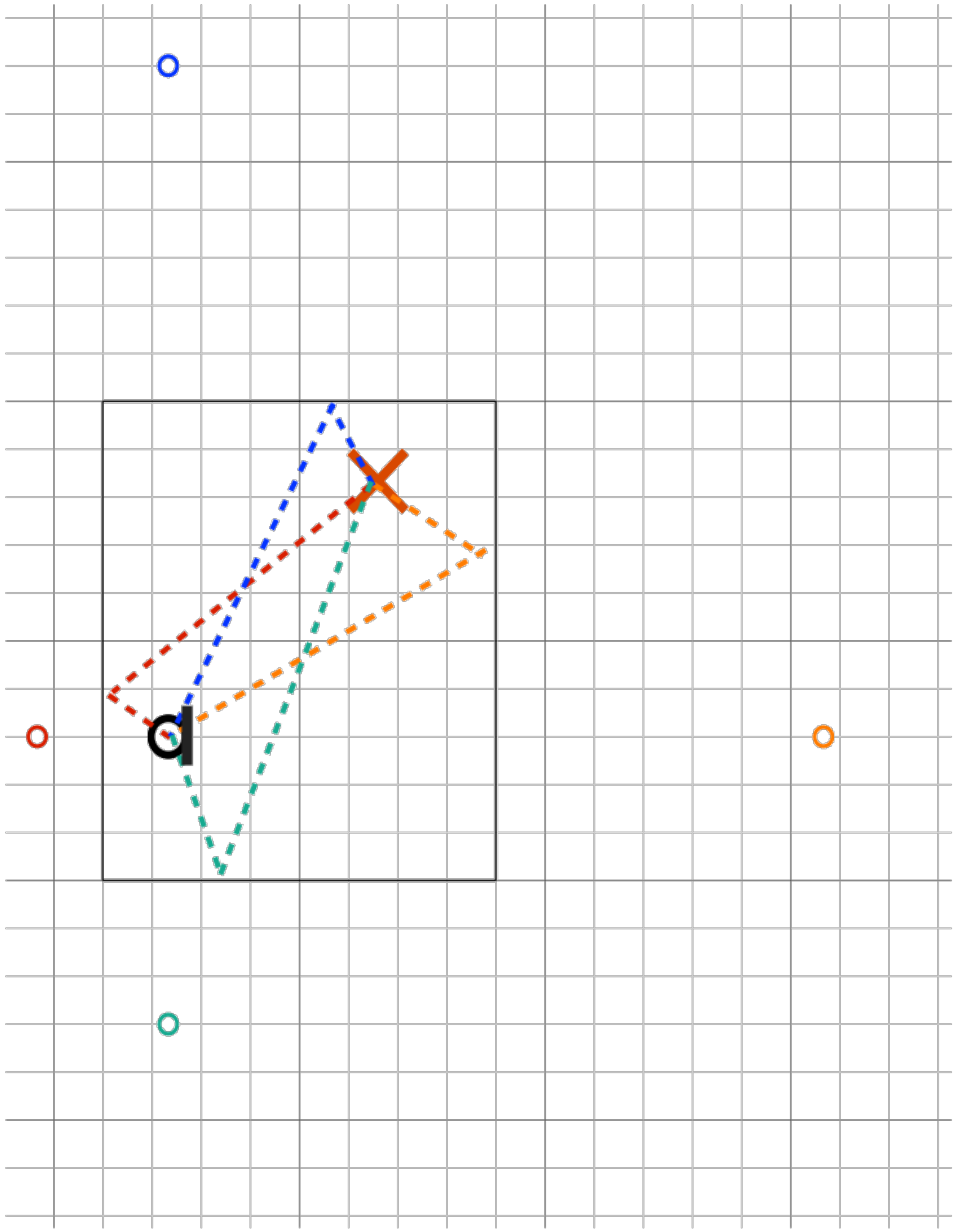
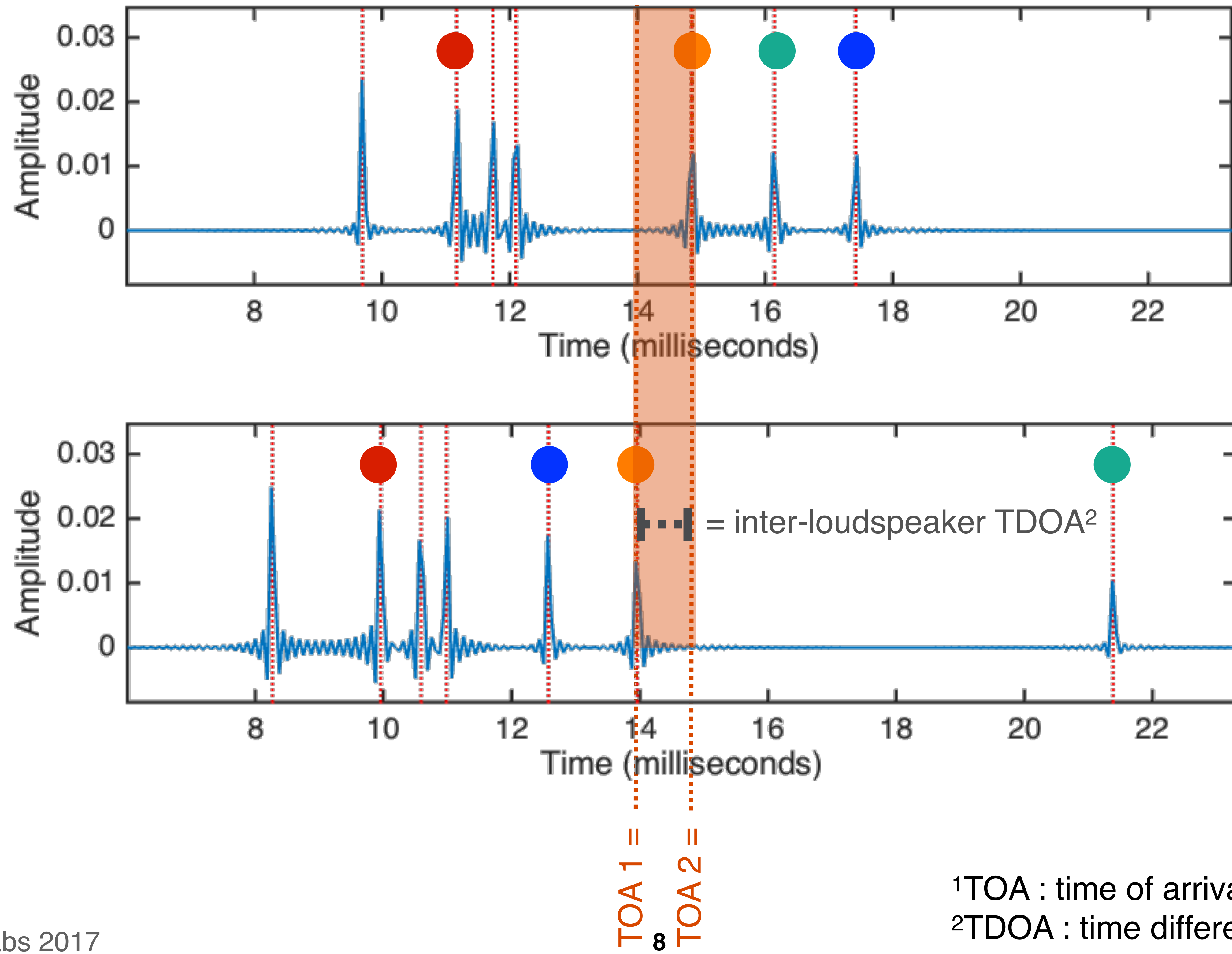


Image Model





Full TOA[1] Ambiguity Problem



[1]TOA : time of arrival

[2]TDOA : time difference of arrival

Input / Output

$$h_j(t) = \alpha_{0_j}\delta(t - \tau_{0_j}) + \sum_{r=1}^{R} \alpha_{rj}\delta(t - \tau_{rj}) + \eta_j(t)$$

$$\Downarrow$$

$$\{\{\tau_{r_j} : \forall j \in \{1 .. L\}\} : \forall r \in \{0 .. R\}\}$$

$j \in \{1..L\}$: loudspeaker index
$r \in \{0\} \cup \{1..R\}$: reflector index

α_{rj} : attenuation coefficient
$\eta_j(t)$: noise

τ_{rj} : propagation time
δ : delta function

AUDIO LABS

Existing Solutions

- Graph [1], range [2] or Euclidean distance matrix - based [3]
- Graph-theory formulation [1, 2, *(3)*]
- Feasibility problem
  - Combinatorial approach [1, 2, 3]
  - Problems on real data [1, 2, 3]
- Reliance on TDOAs [1] instead of TOAs [2, 3]
  - Very tight margins for error for TDOAs
  - High precision / sampling frequency needed

[1] Scheuing et al (2008-2013) [2] Venkateswaran (2012) [3] Dokmanic et al (2013 - 2016)

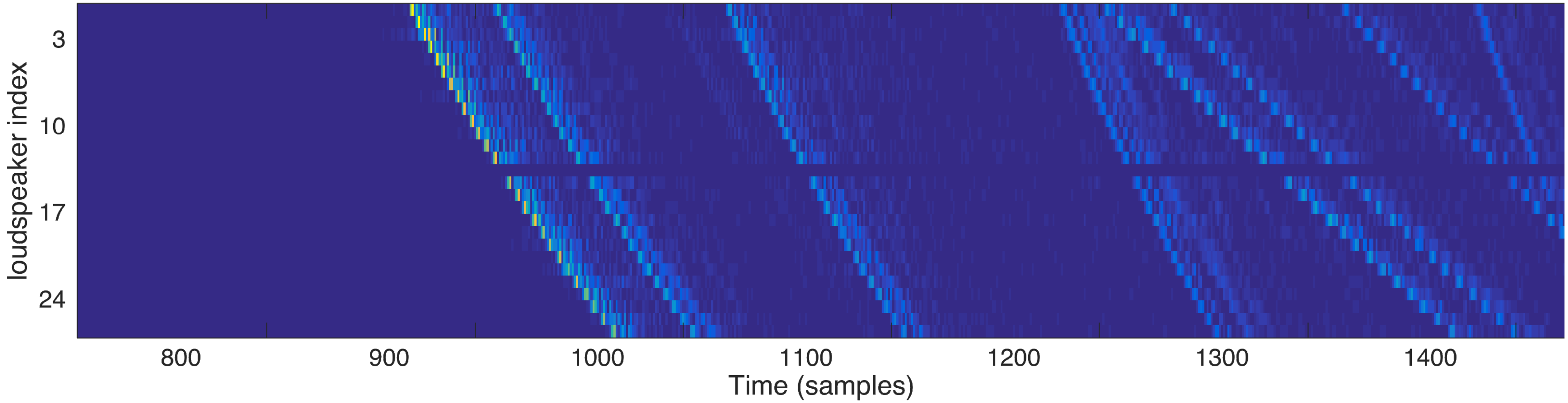
Proposed Solution

I. Stack Room Impulse Responses (RIRs)

II. Detect Lines with the Linear Radon Transform (LRT)

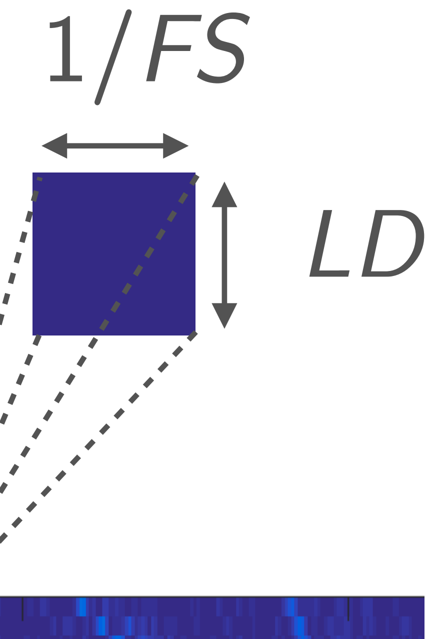
III. Map Lines to TOA Sets

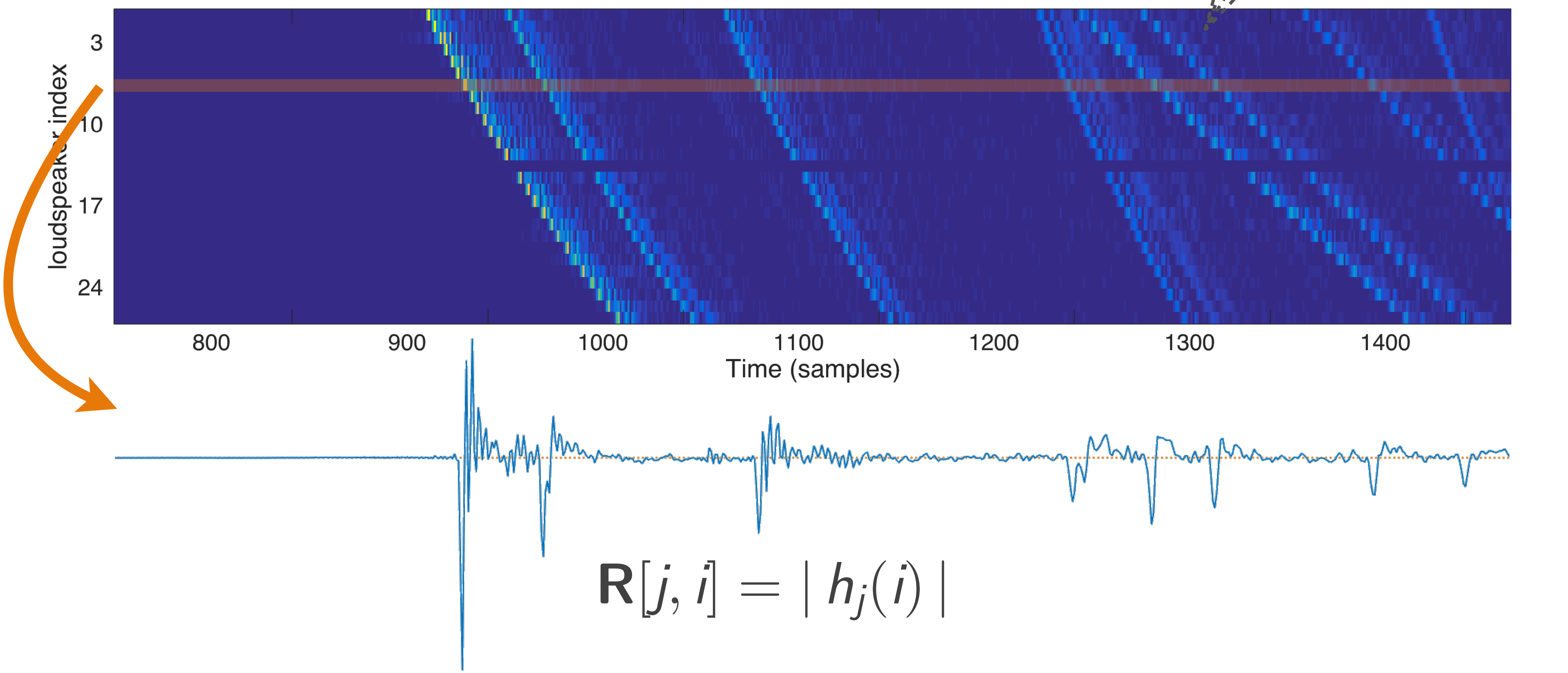
I - RIR Stacking



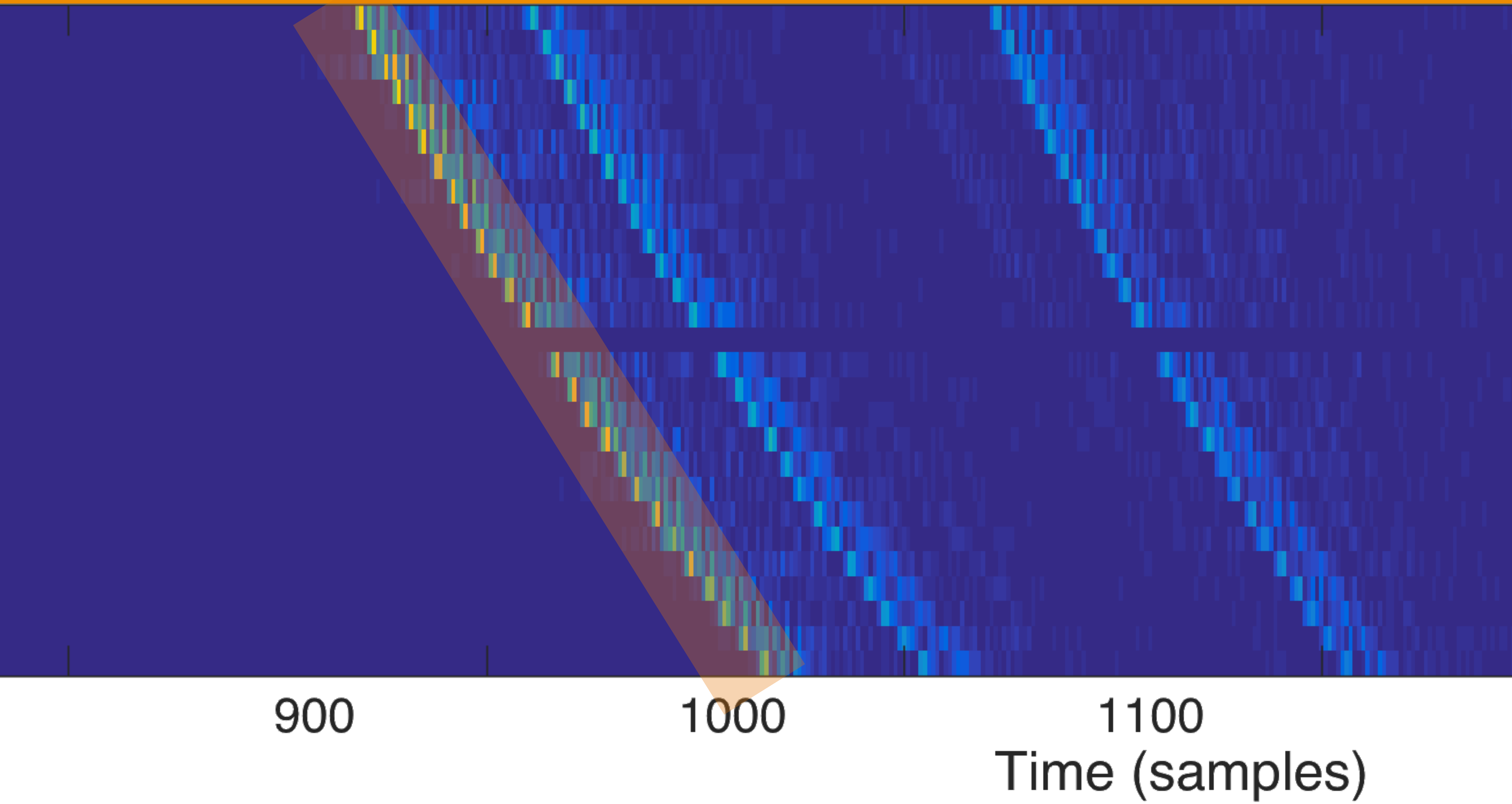
$$\mathbf{R}[j, i] = |\, h_j(i) \,|$$

I - RIR Stacking



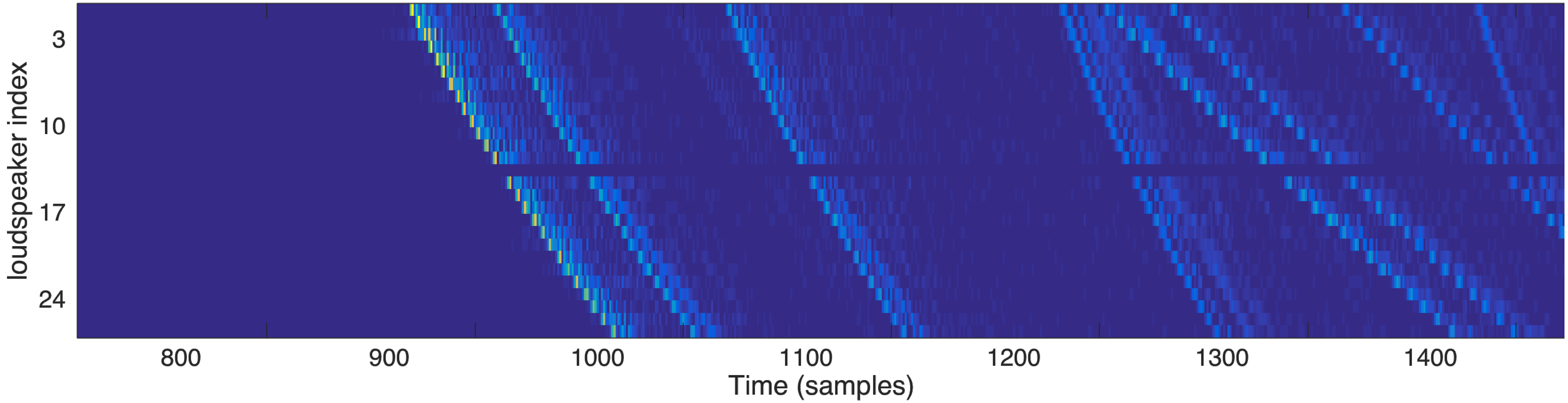


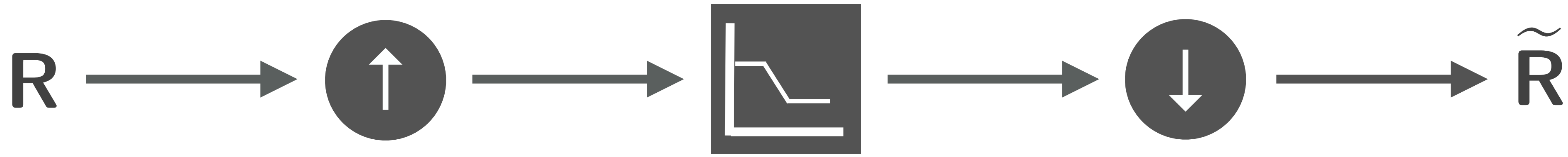
$$\mathbf{R}[j, i] = |\, h_j(i) \,|$$



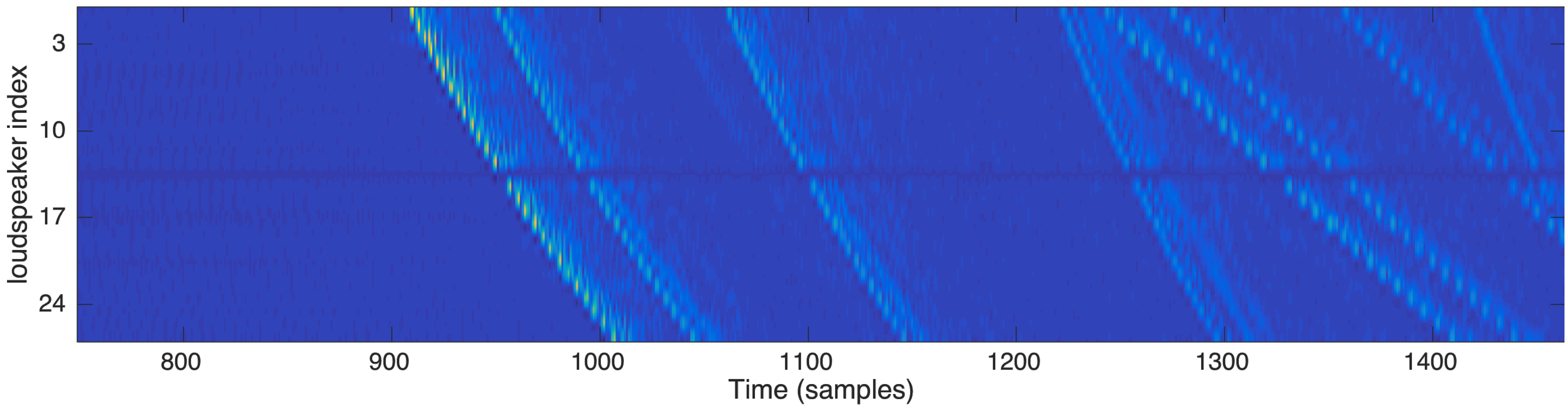
$$\mathbf{R}[j, i] = |\, h_j(i) \,|$$

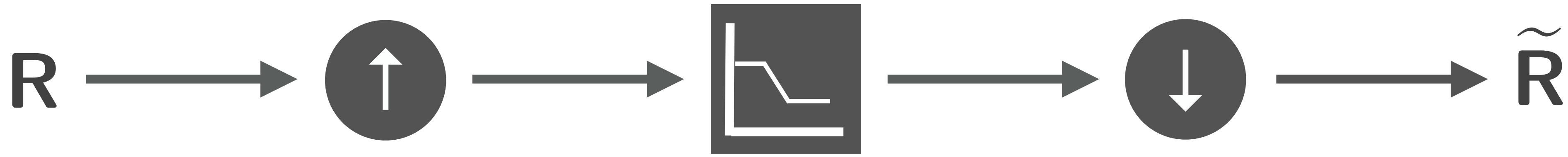
II - RIR Stack Resampling



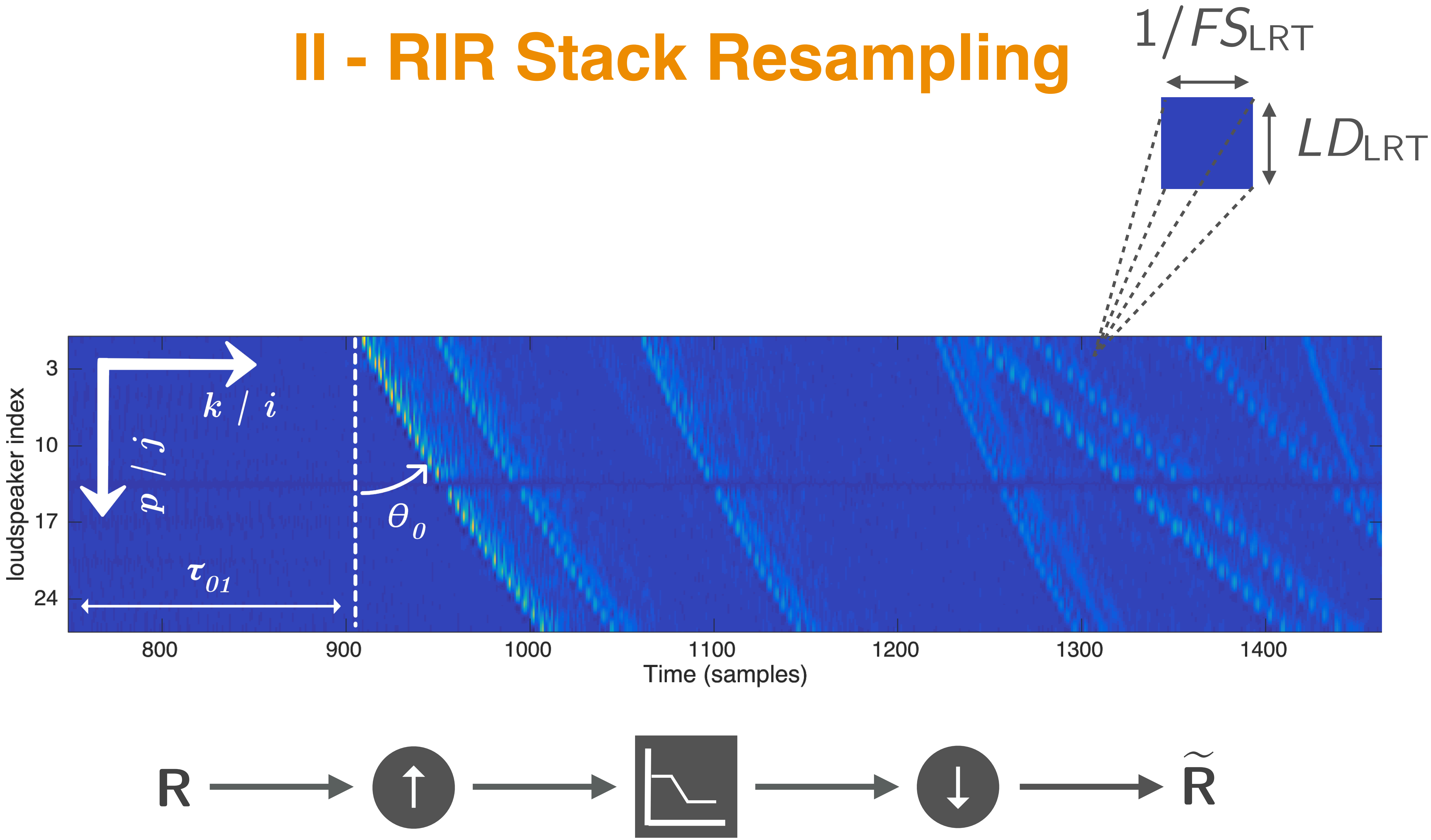


II - RIR Stack Resampling

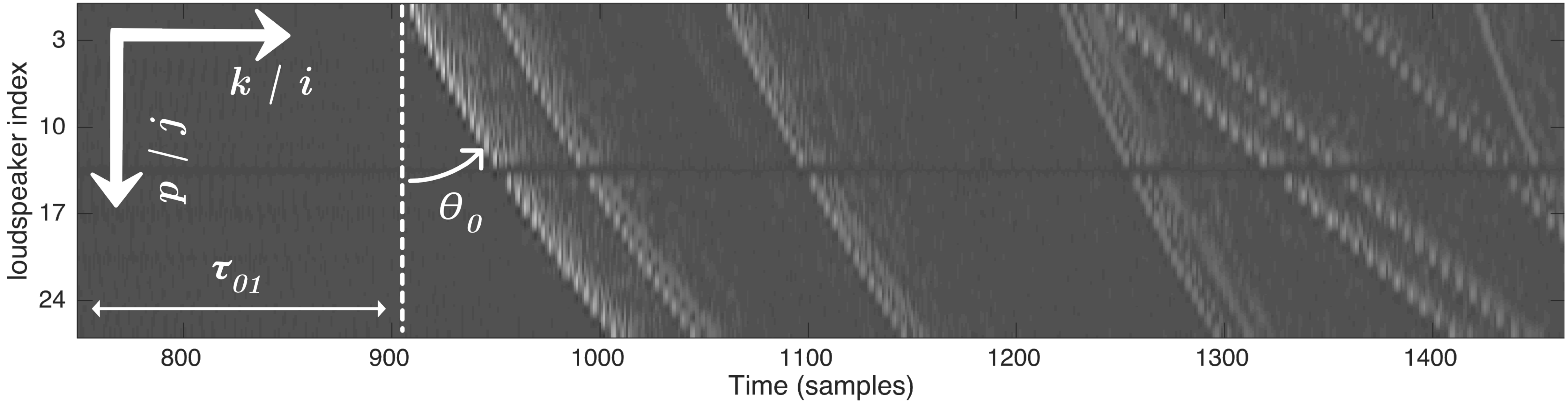




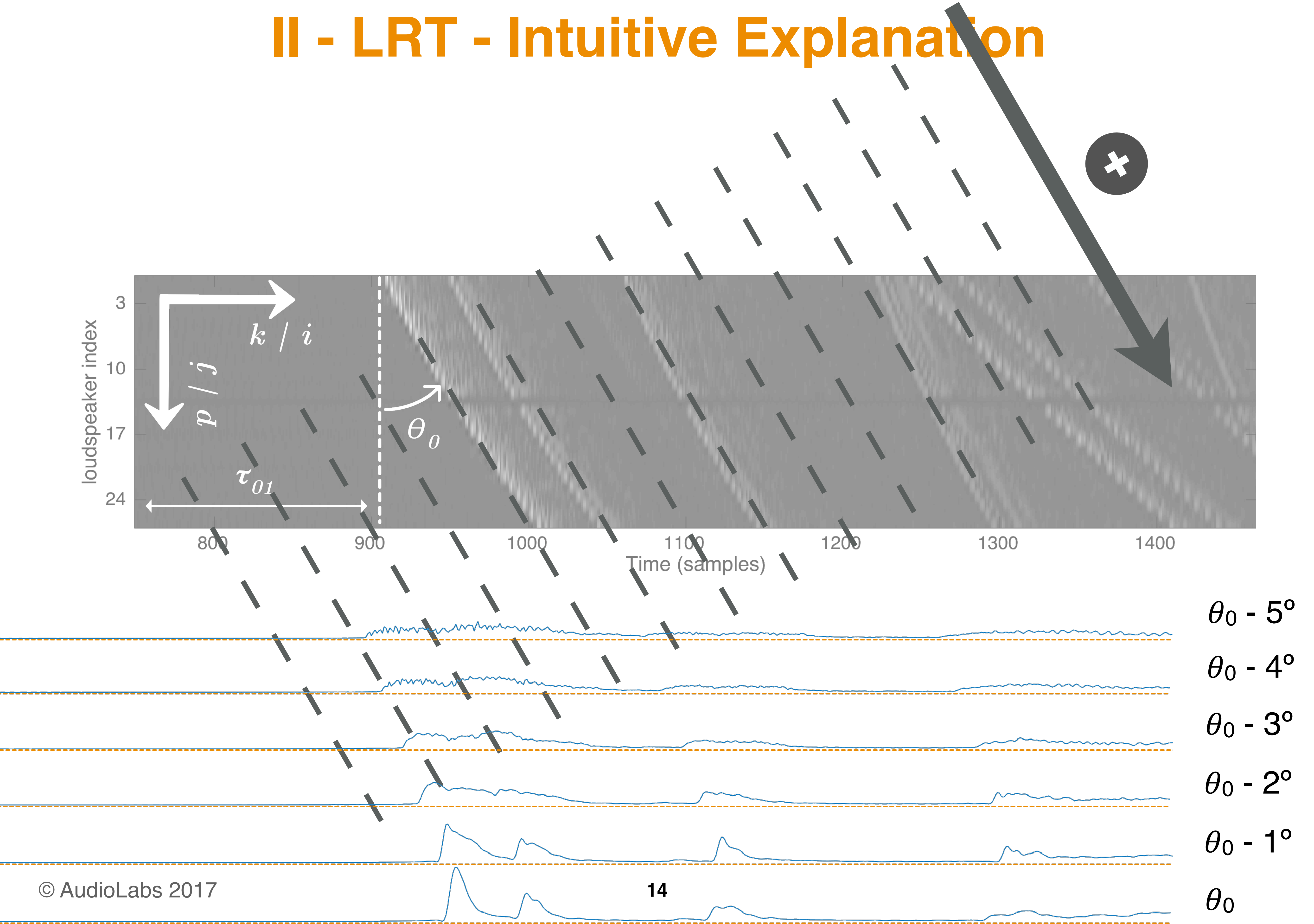
II - RIR Stack Resampling



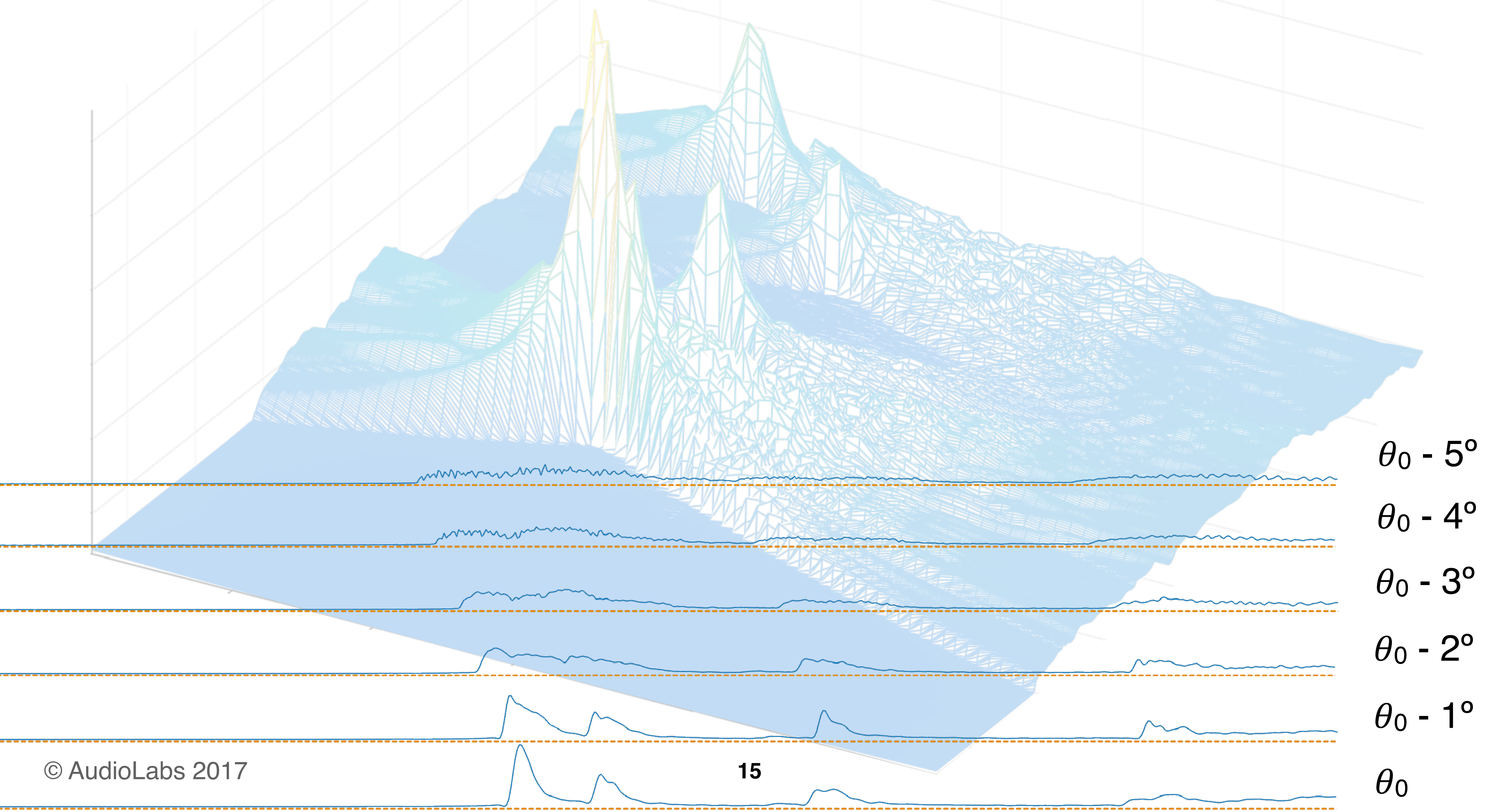
II - LRT - Intuitive Explanation



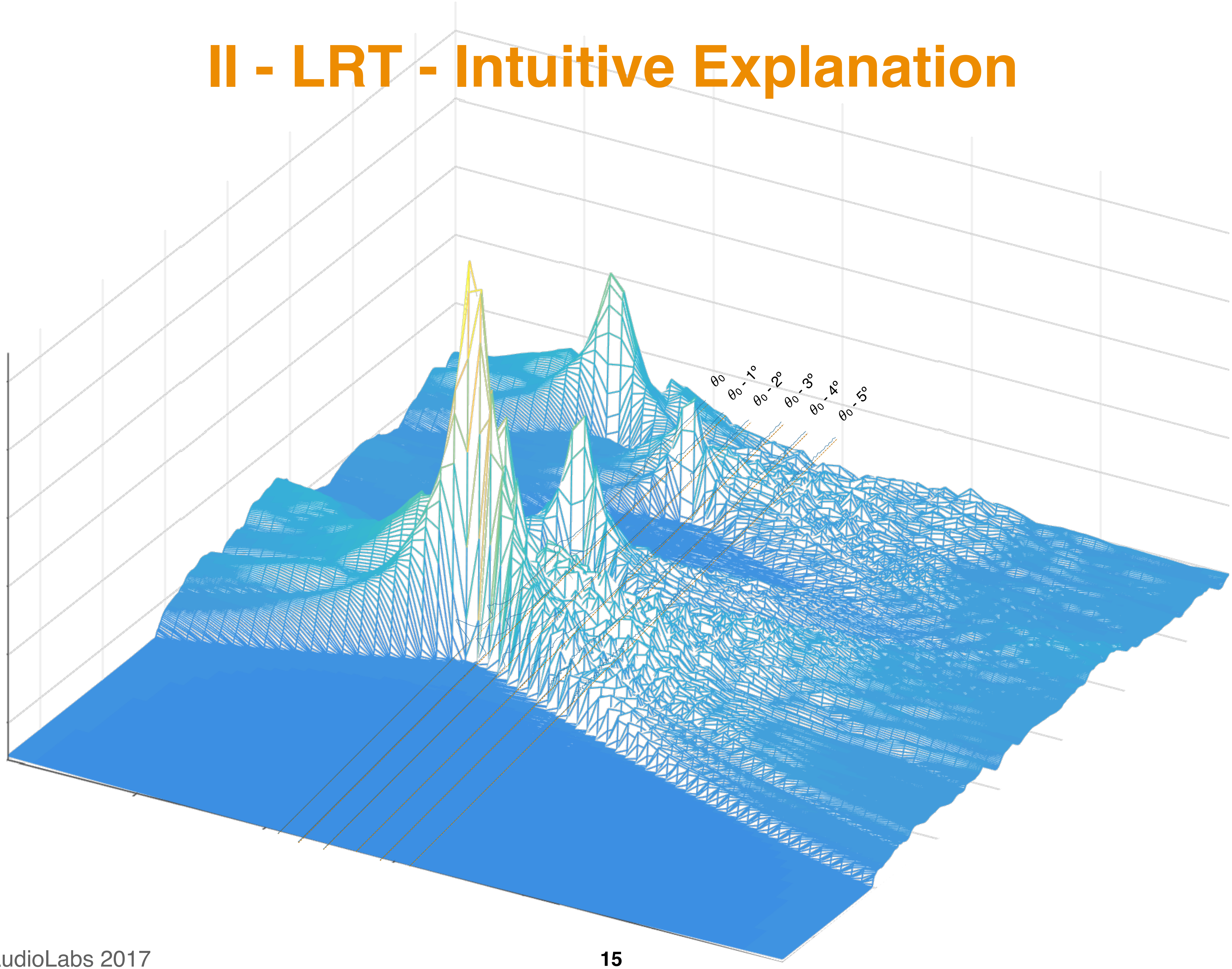
II - LRT - Intuitive Explanation



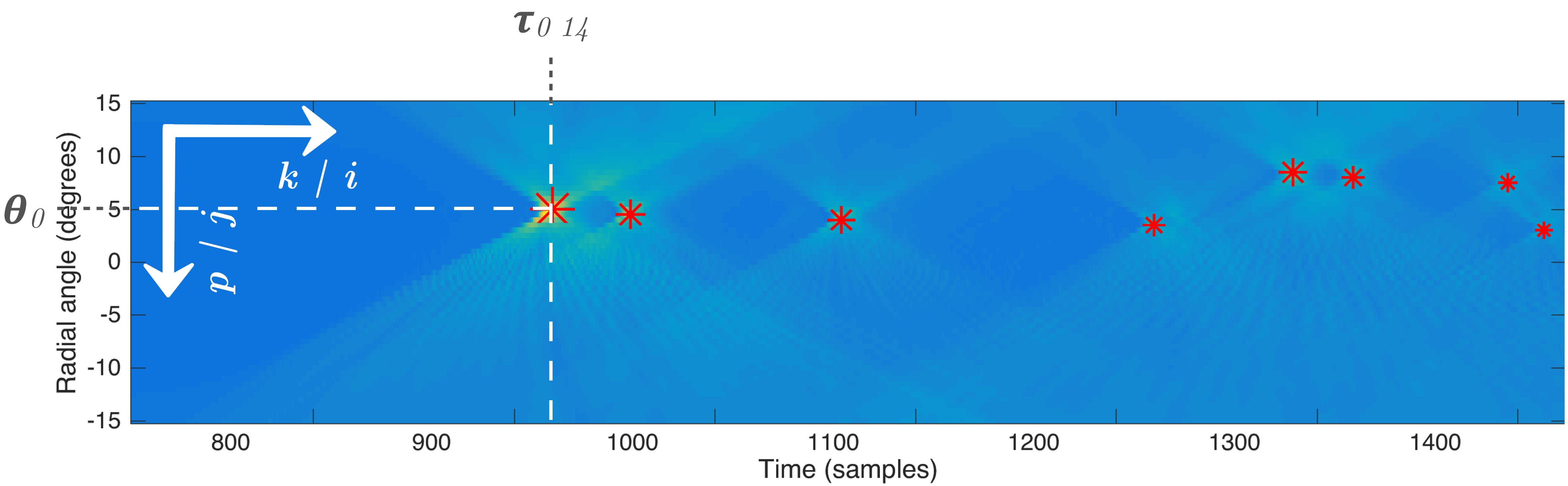
II - LRT - Intuitive Explanation



II - LRT - Intuitive Explanation

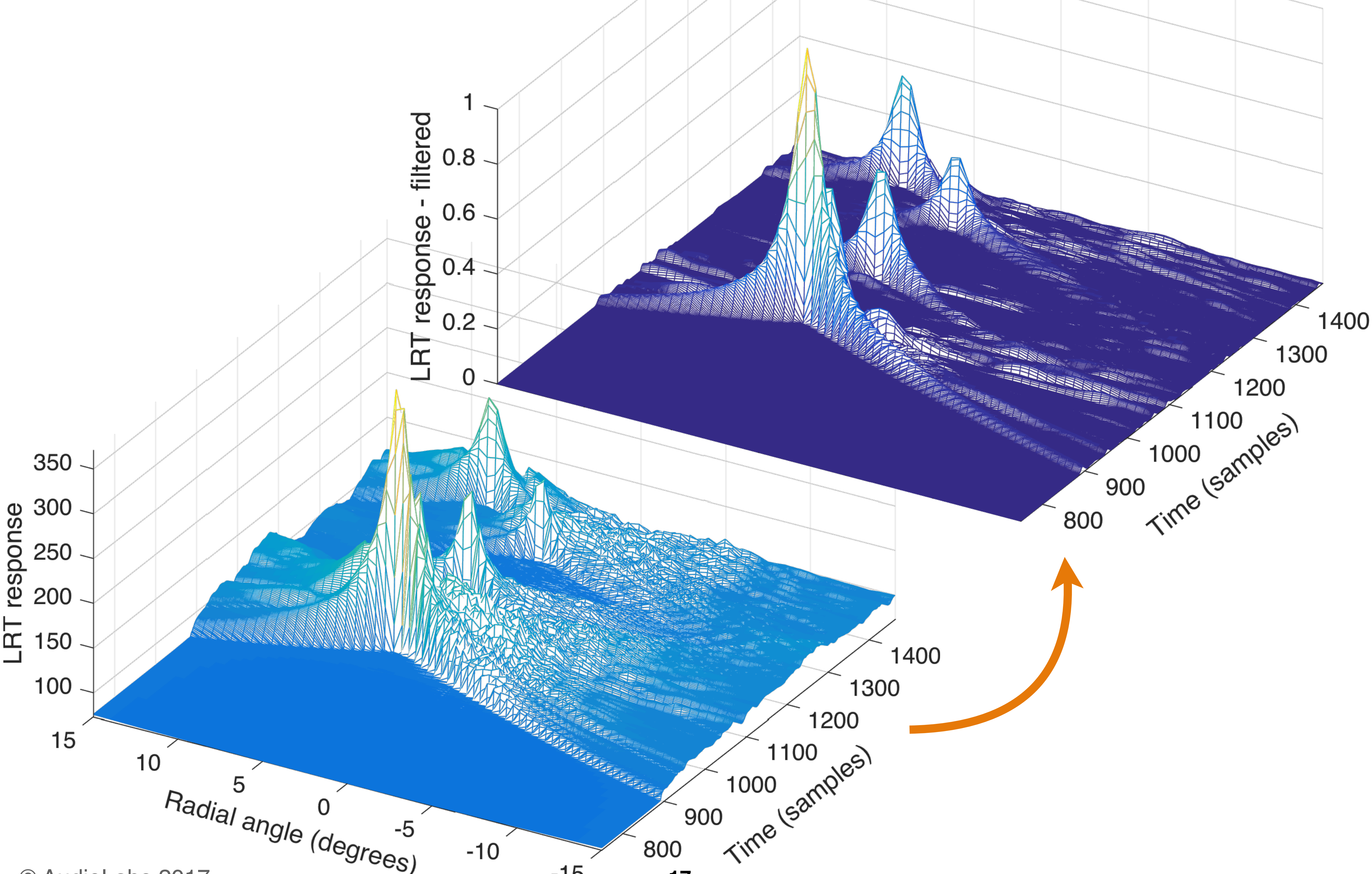


II - LRT - Initial Transform

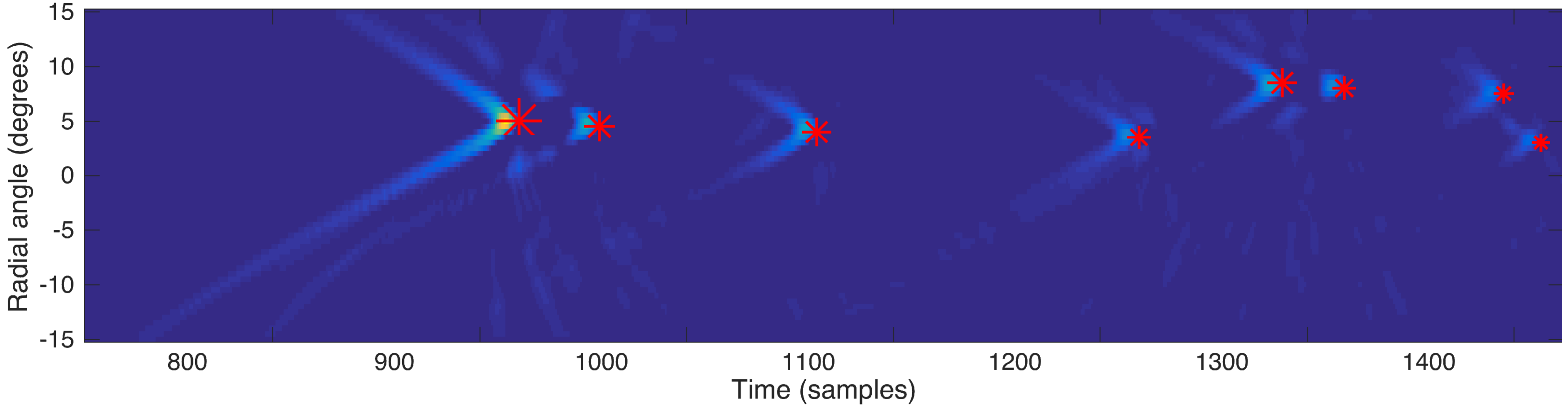


$$\mathbf{T}[j,k] = \sum_{p} \widetilde{\mathbf{R}}[p, k + (p - \widetilde{L}/2) \cdot \tan(\theta_j))]$$

II - LRT - Topography



II - LRT - Filtered Transform

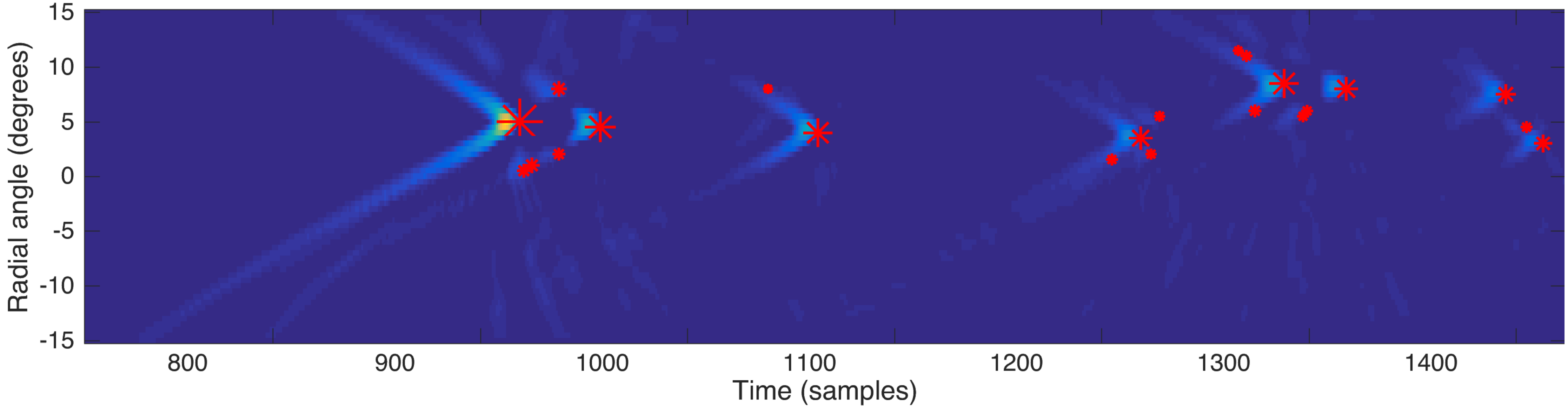


$$\mathbf{T} * \begin{pmatrix} -1 & \cdots & -1 & 0 & 1 & \cdots & 1 \\ -1 & \cdots & -1 & 0 & 1 & \cdots & 1 \\ -1 & \cdots & -1 & 0 & 1 & \cdots & 1 \end{pmatrix}, \quad f(v) = \max(0, v)$$

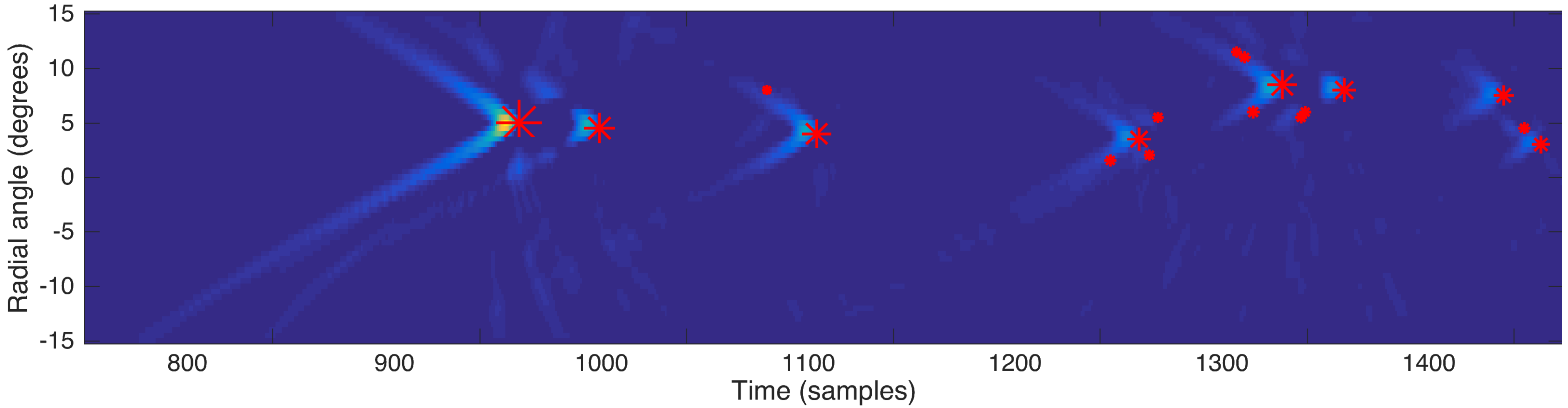
II - Peak Response Detection

- Basic 2D peak picking
- Peak response detection constraints:
  - No peaks earlier than biggest peak
  - No line intersects direct-sound line
  - Other constraints

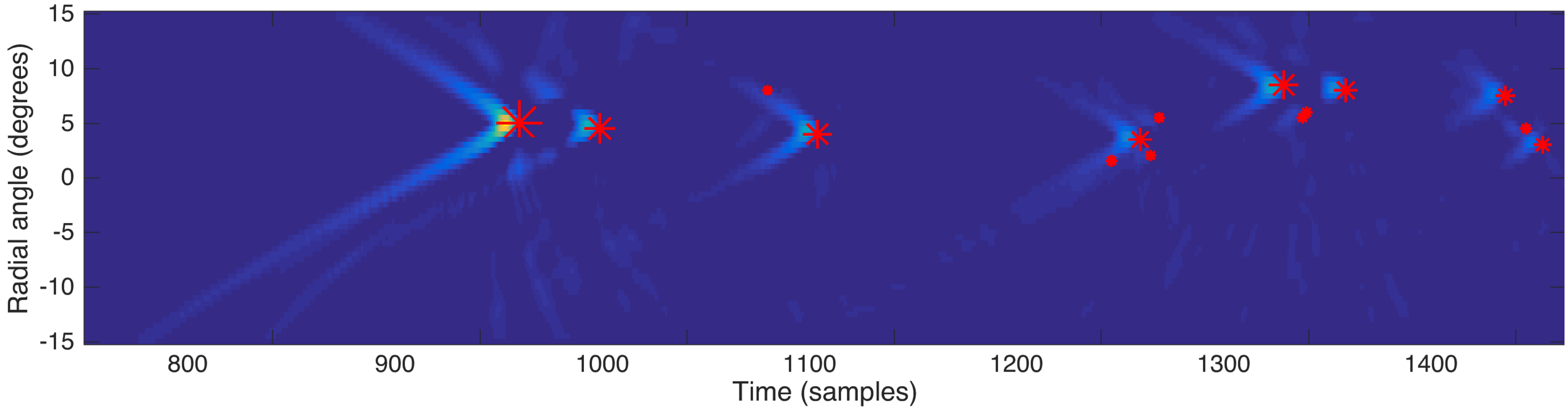
II - LRT - Neighborhood Suppression



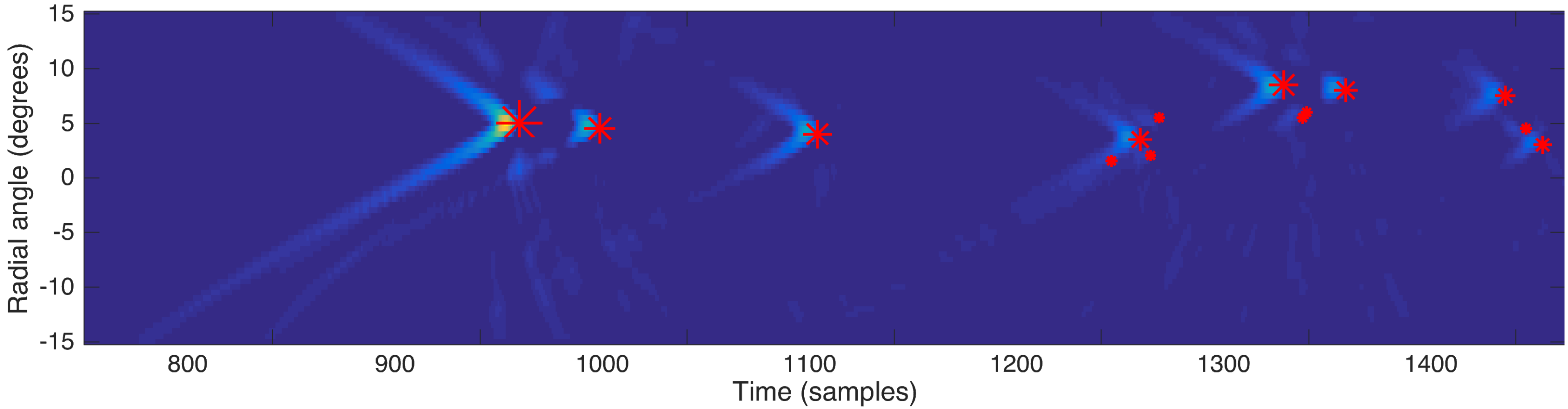
II - LRT - Neighborhood Suppression



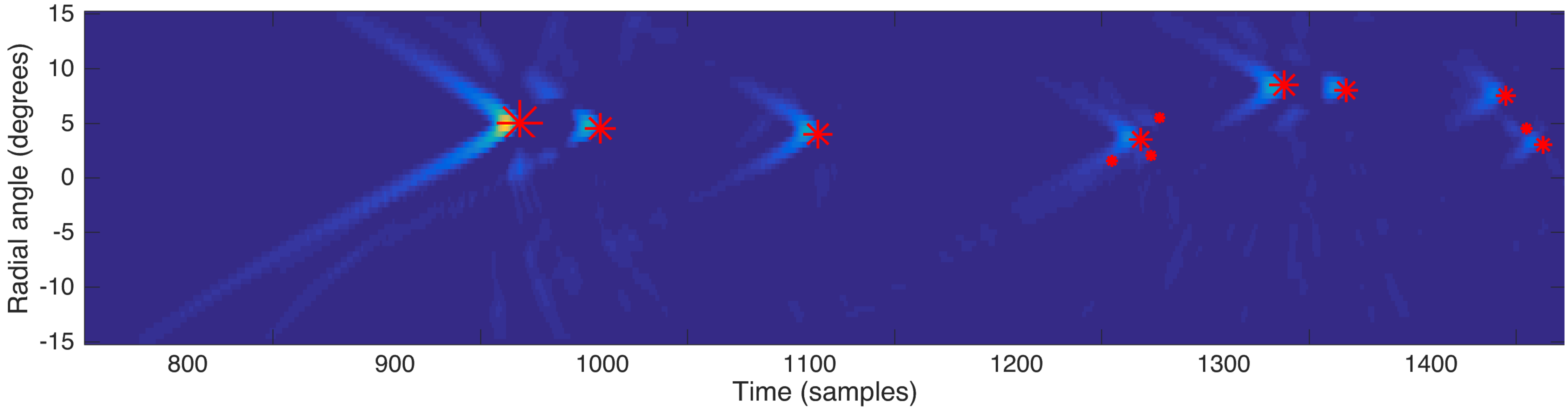
II - LRT - Neighborhood Suppression



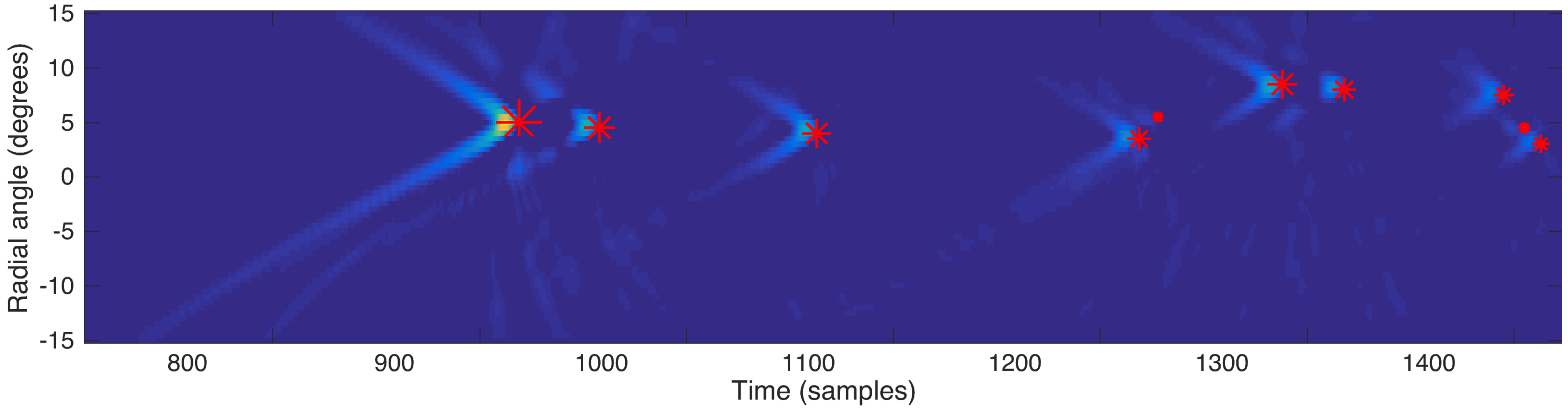
II - LRT - Neighborhood Suppression



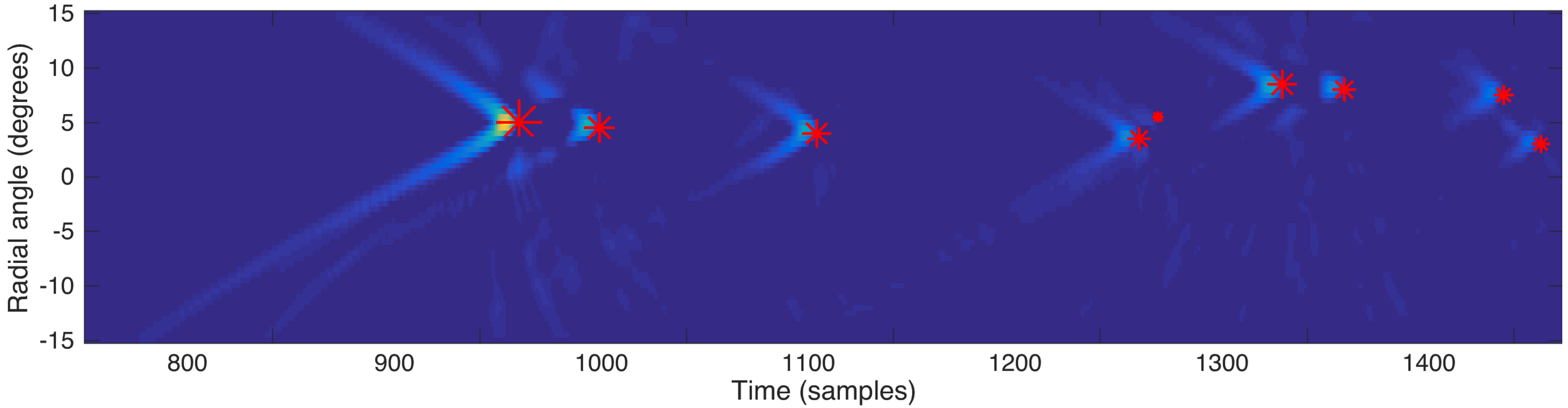
II - LRT - Neighborhood Suppression



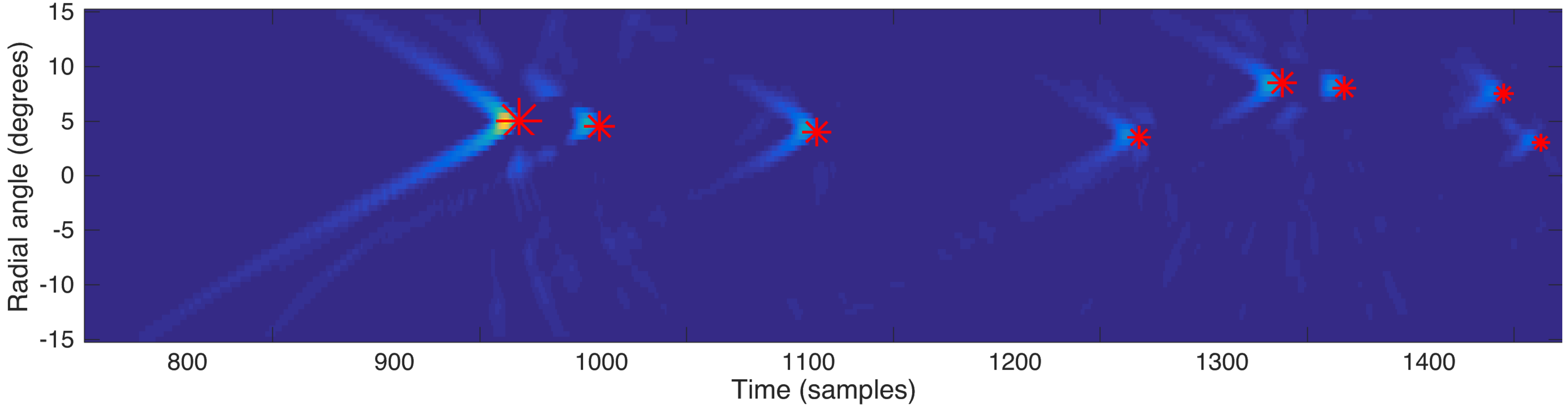
II - LRT - Neighborhood Suppression



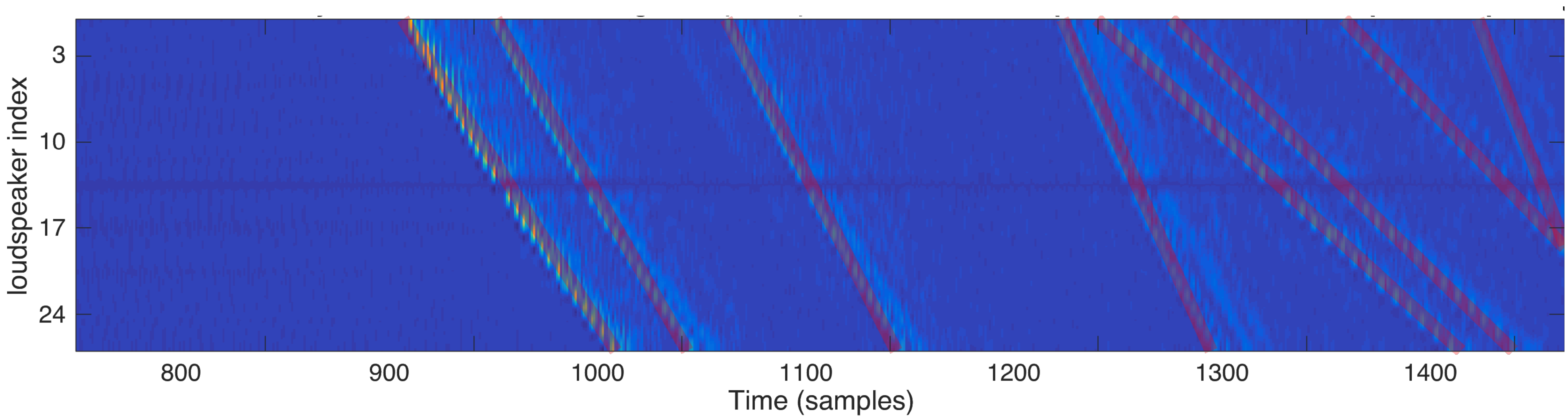
II - LRT - Neighborhood Suppression



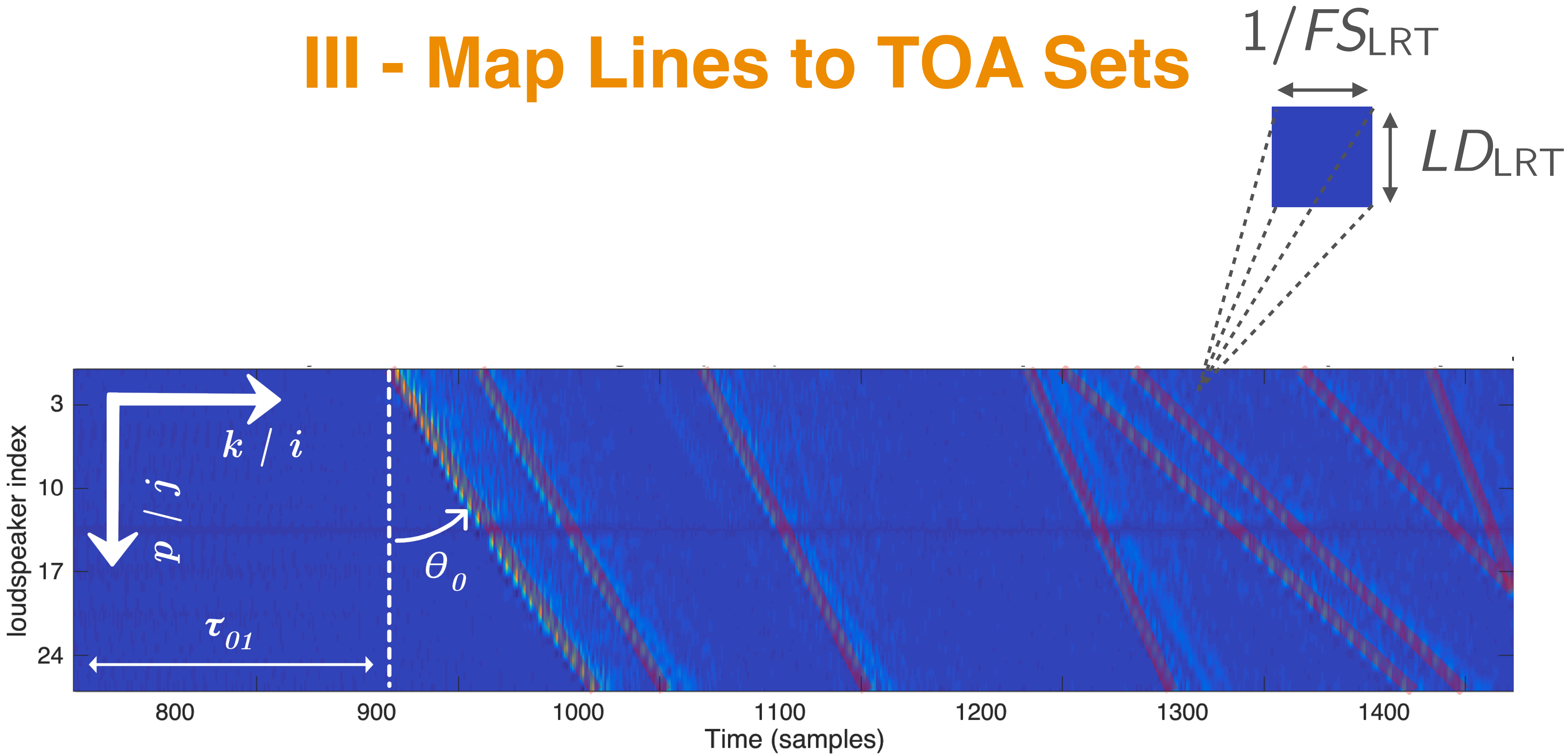
II - LRT - Neighborhood Suppression



II - Final Lines

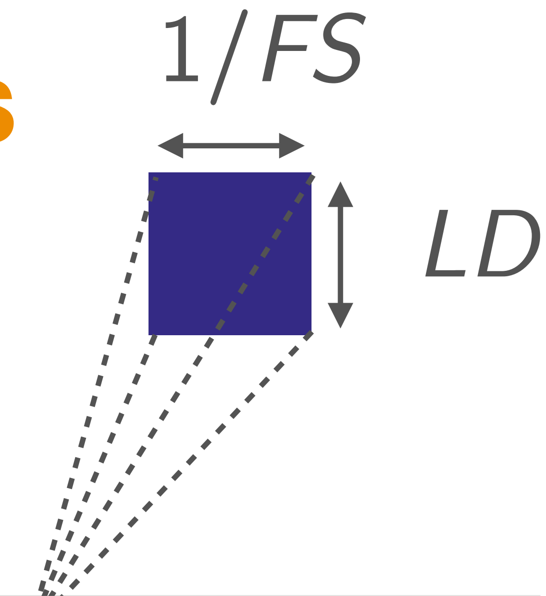


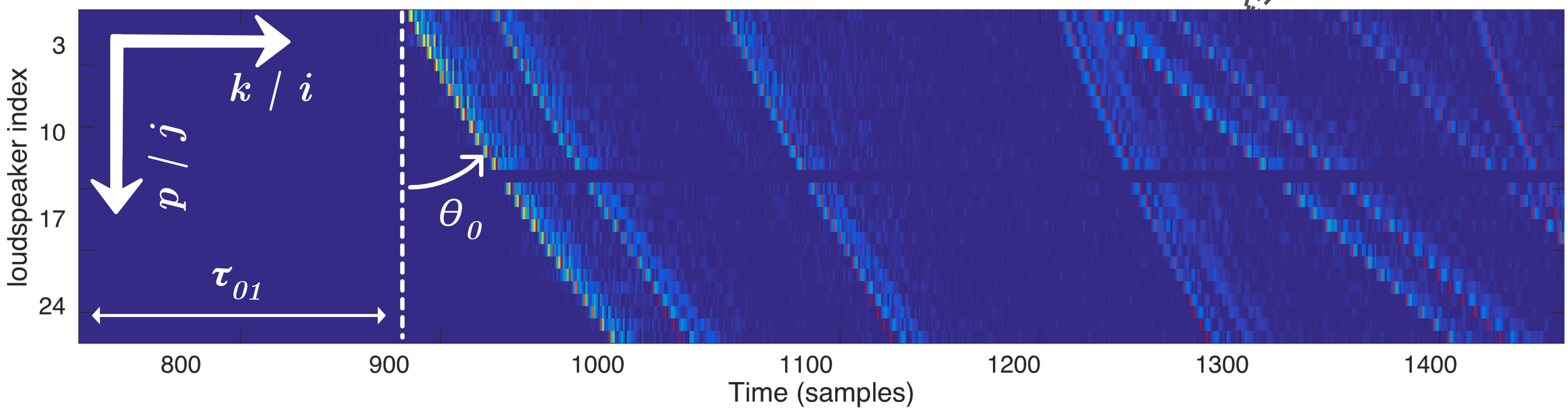
III - Map Lines to TOA Sets



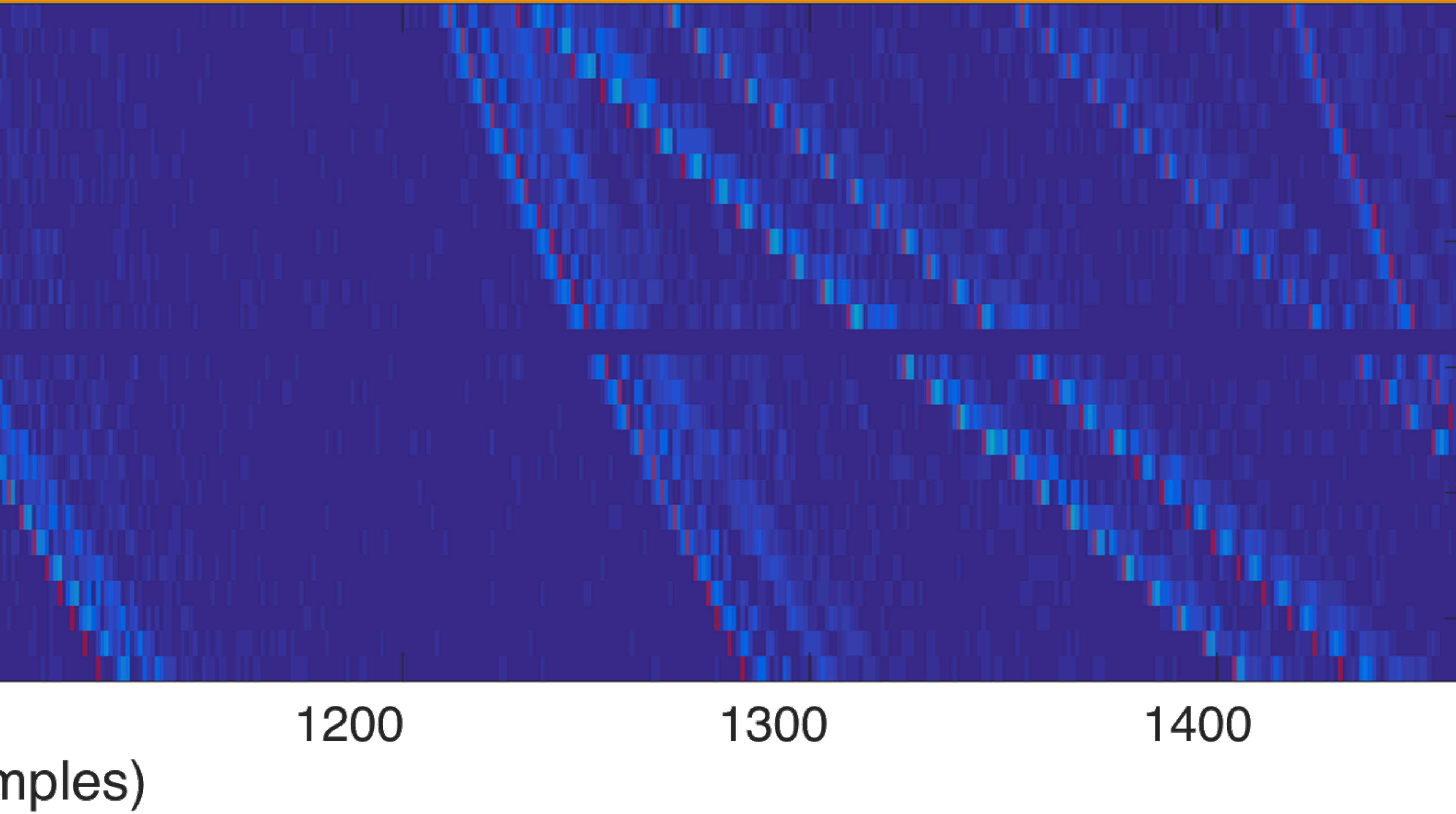
$$\tau_{rl} = (FS_{\text{LRT}} \cdot LD_{\text{LRT}} \cdot \tan(\mathcal{A}(j_r))/(FS^2 \cdot LD)) \cdot (l-1) + k_r/FS_{\text{LRT}}$$

III - Map Lines to TOA Sets





$$\tau_{rl} = (FS_{\mathrm{LRT}} \cdot LD_{\mathrm{LRT}} \cdot \tan(\mathcal{A}(j_r))/(FS^2 \cdot LD)) \cdot (l - 1) + k_r/FS_{\mathrm{LRT}}$$



$$\tau_{rl} = (FS_{\text{LRT}} \cdot LD_{\text{LRT}} \cdot \tan(\mathcal{A}(j_r))/(FS^2 \cdot LD)) \cdot (l-1) + k_r/FS_{\text{LRT}}$$

AUDIO LABS

Performance Evaluation - Metrics

- Three metrics for TOA sets:
  - True Positive Rate (TPR)
  - False Discovery Rate (FDR)
  - Root Mean Square Error (RMSE) of TOAs
- One-to-one match with RMSE < 0.5ms is correct

Performance Evaluation - Setups

- 6 simulated setups: 9 mics and 16 loudspeakers each
- 1 real setup: 8 mics and 27 loudspeakers each

Performance Evaluation - Results

		direct-sound		**first-order reflections**		**second-order reflections**	
Metrics / Setups	FDR (%)	TPR (%)	RMSE (μs)	TPR (%)	RMSE (μs)	TPR (%)	RMSE (μs)
Sim. data	0.63	100	104.8	92.6	66.5	37.7	59.9
Real data	5.83	100	116.1	83.3	138.7	32.0	166.3

Conclusion

- Robust echo labeling
- Works exclusively on linear arrays
- High accuracy / performance
- Efficient and intuitive

Thank you for your attention

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Fraunhofer IIS