Hunting Echoes
for
Auditory Scene Analysis

Diego Di Carlo • 27.05.2019

Supervised by
Nancy BERTIN, Antoine DELEFORGE

Outline

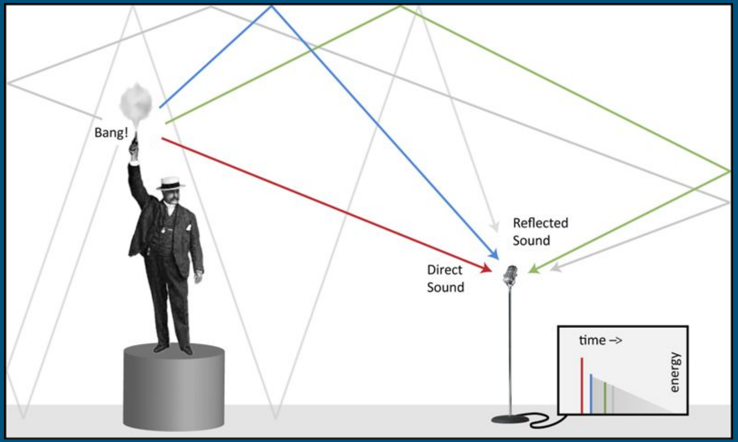
1. What are acoustic echoes?
2. If you know them, we can (1st year)
    a. *Sound Source Separation and* **SEPARAKE**
    Presented at ICASSP18
    b. *Sound Source Localization and* **MIRAGE**
    Presented at ICASSP19
3. How to know them? (2nd year)
    a. Continuous Dictionary and **BRAIRE**
    Work in Progress
    b. Learning-based approach and MIRAGE
    Presented at ICASSP19
4. Applications
    a. *Honda Haru*

ACOUSTIC ECHOES

(Room) Impulse Response: the linear filtering effect due to the propagation of sound from a source to a microphone in a room.

It accounts for ...

- … the geometry of the audio scene:
  - Room shape and size
  - Source position
  - Microphone position
  - … other objects (e.g. furnitures) sizes and shape.
- … the acoustic properties of the audio scene:
  - Wall materials
  - … Type of other objects (e.g. furnitures materials)



ACOUSTIC ECHOES

RIR can be subdivided in:

- Direct (or anechoic) path
- Early reflections (**Echoes**)
- Late Reverberation

For some audio inverse problems, the sound propagation is typically…

- … ignored *[Le Roux et al. 2015, DC et al. 2017]*;
- … assumed as a single anechoic path *[Rickard 2007]*;
- … modelled entirely *[Ozerov et al. 2010, Nugraha et al. 2016]*;
- … assumed as late reverberation *[Leglaive et al. 2016]*.

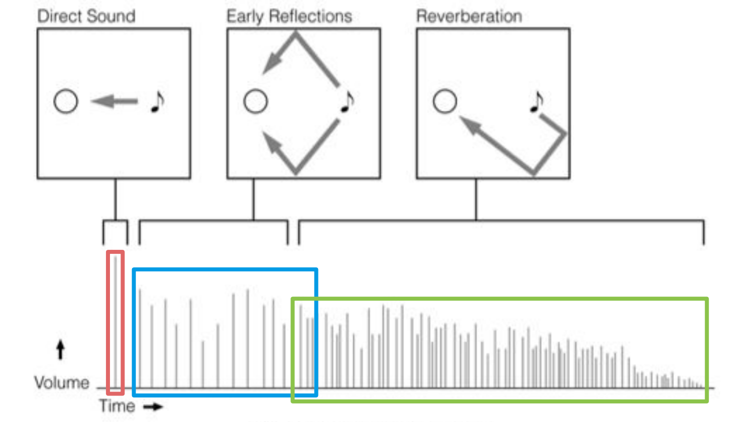


Figure 1: Elements of Reveberation

If we know them?

Reverberation has a detrimentally affects typical **Audio Inverse Problem** algorithm.

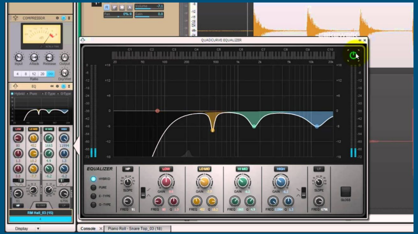
Can Echoes help?

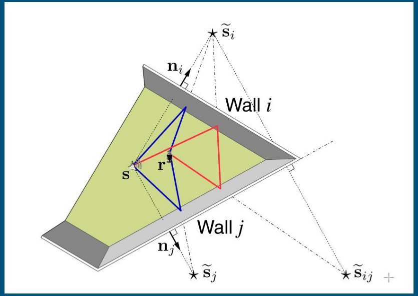
Recents **echo-aware** methods showed that the knowledge of early echoes increases their performances.

<u>HP : Assume we know echo's coefficients and locations</u>

Inverse Problems with Echo-aware methods

- Speech enhancement
  - Sound Source Localization
    - [Riberio et al. 2010, DC 2019, ...]
  - Sound Source Separation
    - [Scheibler et al. 18, Leglaive et al. 2016]
  - Beamforming
    - [Dockmanic et al. 2015]
- Microphone calibration
  - [Salvati et al. 16, Dockmanic et al. 2013]
- Dereverberation and Room Equalization
  - [Krishnan et al. 2018]
- Room Geometry Estimation
  - [Crocco et al. 2016,Dockmanic et al. 2013, ...]





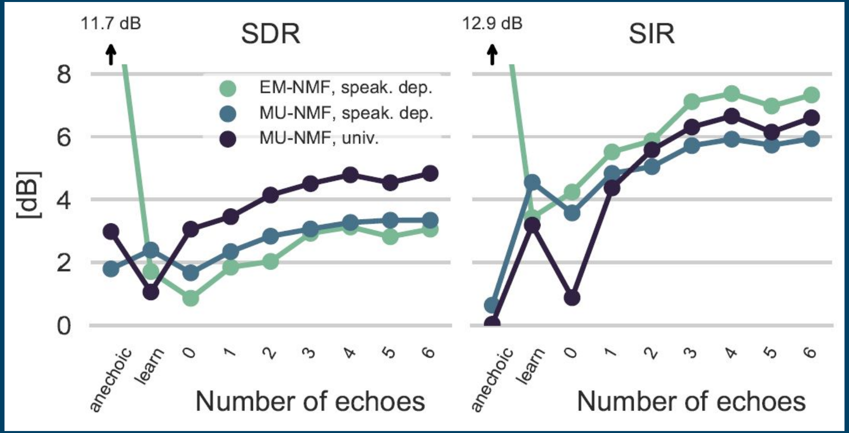
SEPARAKE

Sound source separation with a little help from echoes

Presented at ICASSP 2018

[Robin Scheibler, Diego Di Carlo, Antoine Deleforge, Ivan Dokmanic]

Use transfer function models taking into account early echoes in NMF-based sound source separation methods



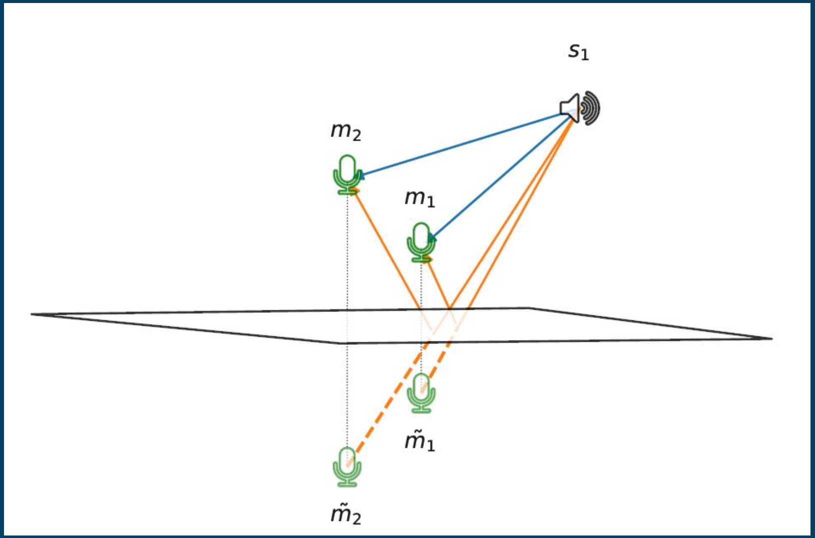
Early echoes are assumed perfectly known in this work

MIRAGE

Microphones array augmentation with echoes

Accepted at ICASSP 2019

[Diego Di Carlo, Antoine Deleforge, Nancy Bertin]

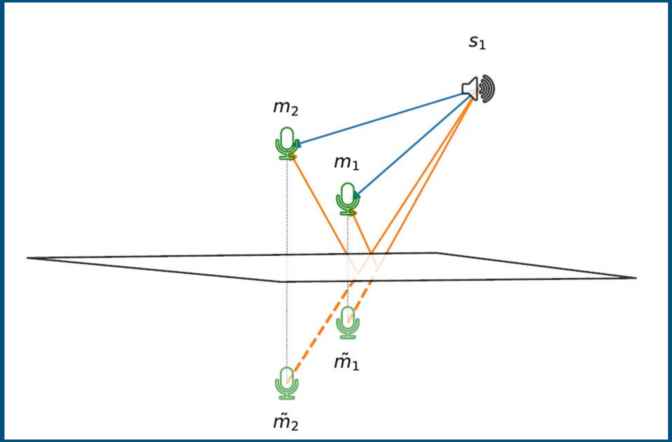


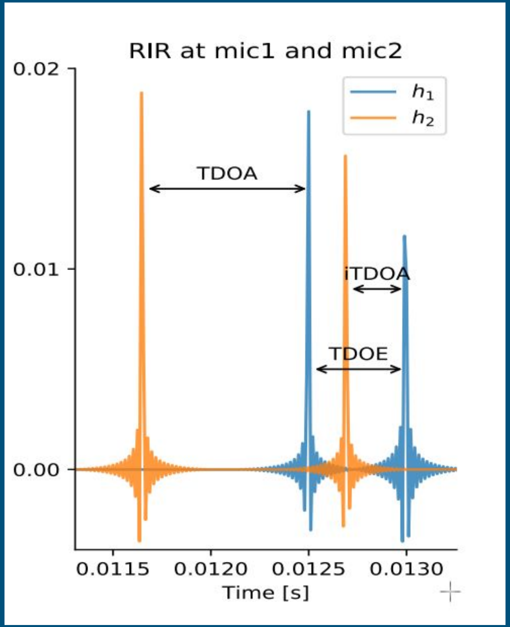
The signal received at m_1 can be seen as the ***sum*** *of anechoic signals received at m_1 and an* ***image microphone $\tilde{m}_1$***

Image source -> Image microphones

Sound Source **Localization** *with a little help from echoes*

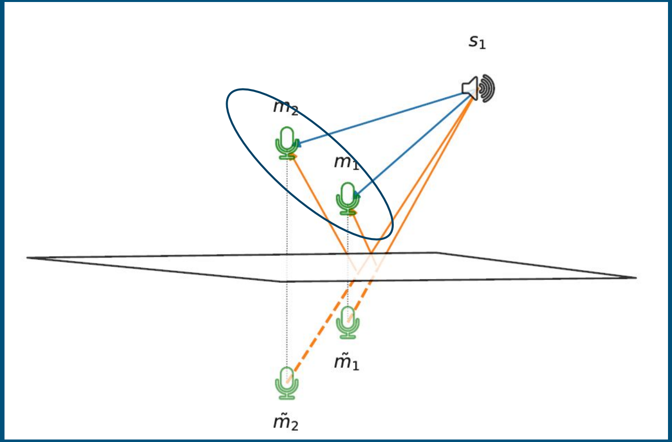
- More microphones… better audio signal processing!
- How to « access » image microphones?

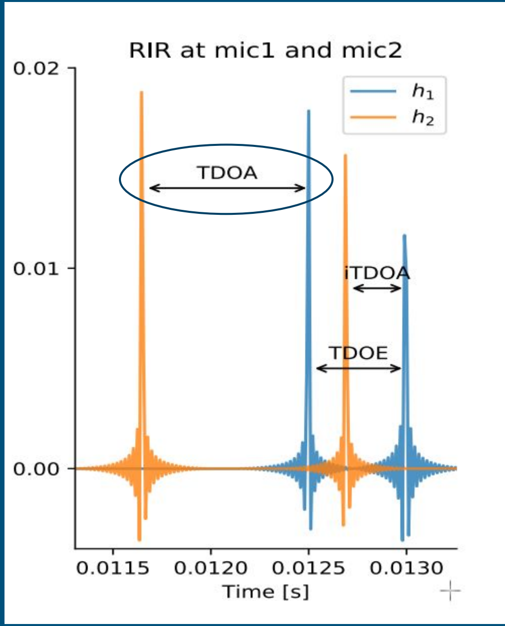




Sound Source **Localization** *with a little help from echoes*

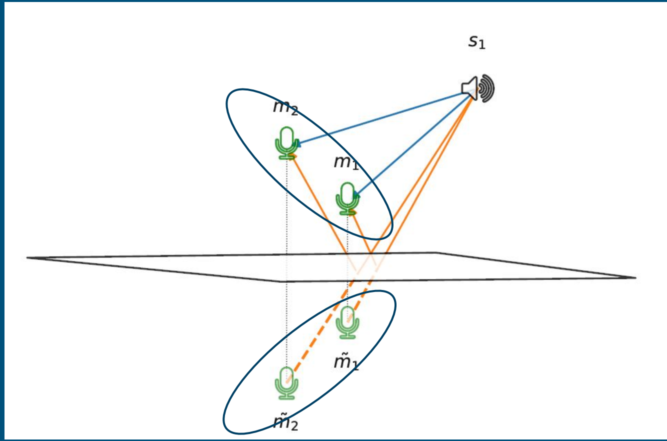
- More microphones… better audio signal processing!
- How to « access » image microphones?

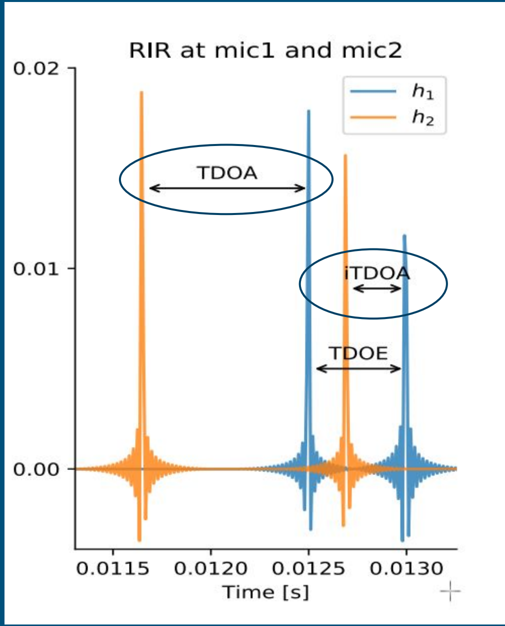




Sound Source **Localization** *with a little help from echoes*

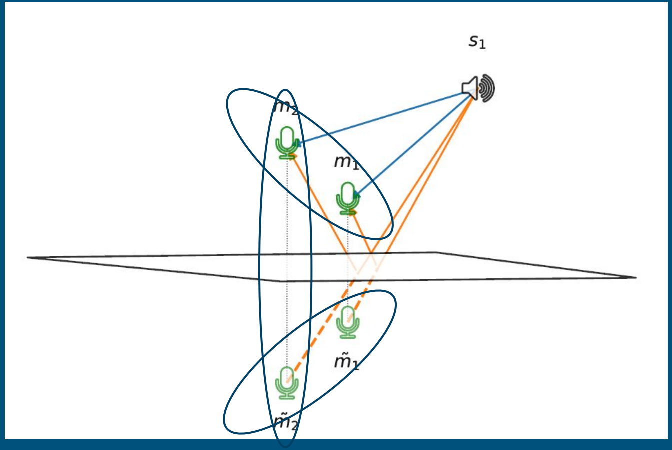
- More microphones… better audio signal processing!
- How to « access » image microphones?

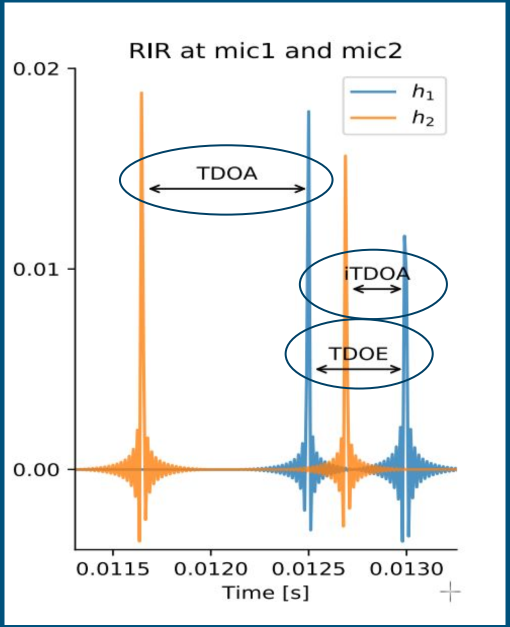




Sound Source **Localization** *with a little help from echoes*

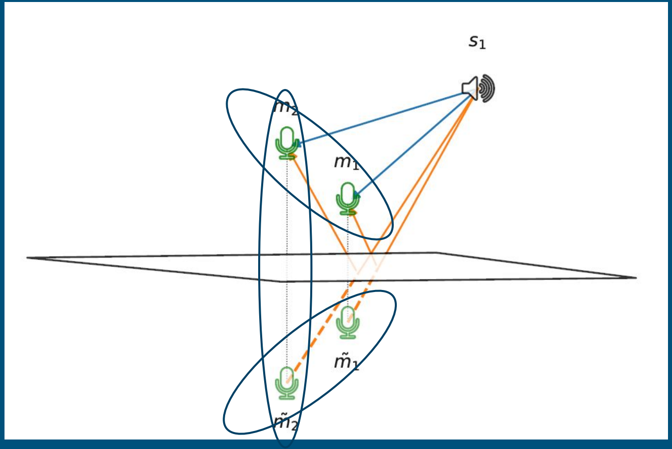
- More microphones… better audio signal processing!
- How to « access » image microphones?

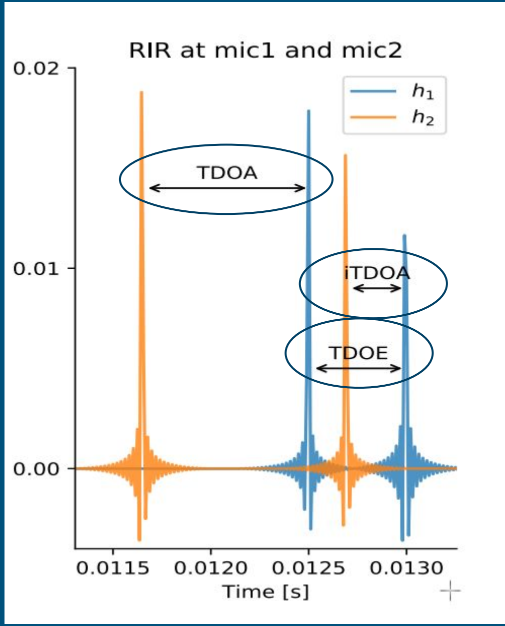




Sound Source **Localization** *with a little help from echoes*

- More microphones… better audio signal processing!
- How to « access » image microphones?





- Each pair in the **augmented array** is associated to impulse response characteristics

Sound Source **Localization** *with a little help from echoes*

- From real numbers to angular spectra

Array Geometry

Pair aggregator (MBSSLocate)

$TDOA_s$

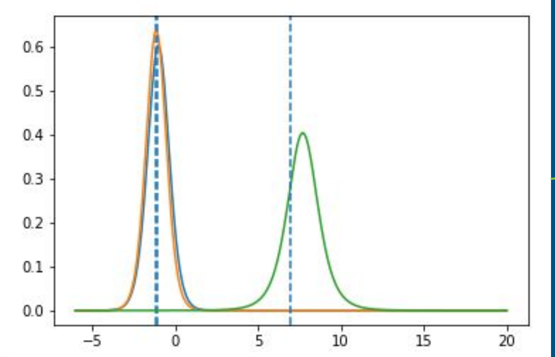
$iTDOA_s$

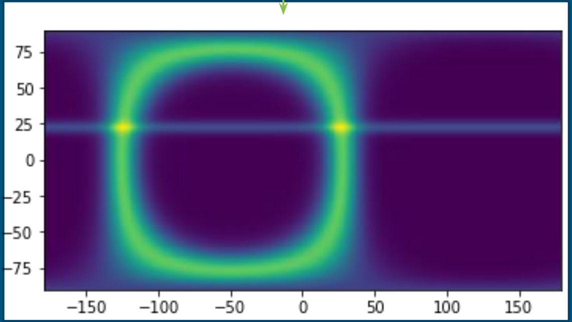
$TDOE_s$

error on the validation set σ^2

μ

Gaussian function





DNN prediction

Local Angular Spectrum

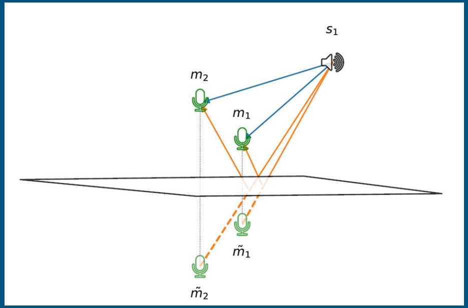
Global Angular Spectrum

Sound Source **Localization** *with a little help from echoes*

- Aggregating (with MBSSLocate) time differences of arrival from multiple microphone pairs enables 2D sound source localization
- The microphone and surface positions are assumed known
- Promising «**impossible localization**» results using clean signals and white noise sources
- Future work:
    - Aggregating multiple pairs
    - Test on real data
    - Perfect symmetries breaks the model

Results on test set
[ICASSP19]

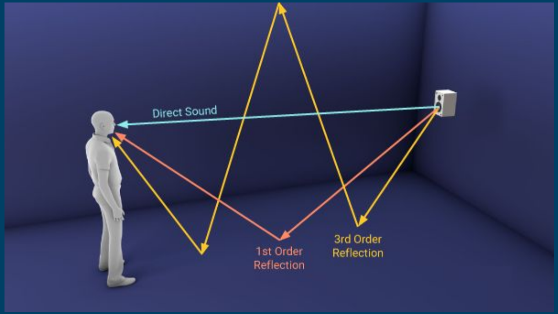
DoA	Input	ACCURACY $< 10^\circ$ θ	ACCURACY $< 10^\circ$ ϕ	ACCURACY $< 20^\circ$ θ	ACCURACY $< 20^\circ$ ϕ
MIRAGE	wn	4.5 (59)	3.9 (71)	6.8 (79)	5.9 (88)
MIRAGE	wn+n	4.4 (18)	5.5 (26)	9.4 (35)	11.1 (66)
MIRAGE	sp	4.6 (45)	4.8 (59)	8.1 (71)	7.2 (83)
MIRAGE	sp+n	5.2 (17)	5.9 (12)	10.7 (38)	12.3 (43)

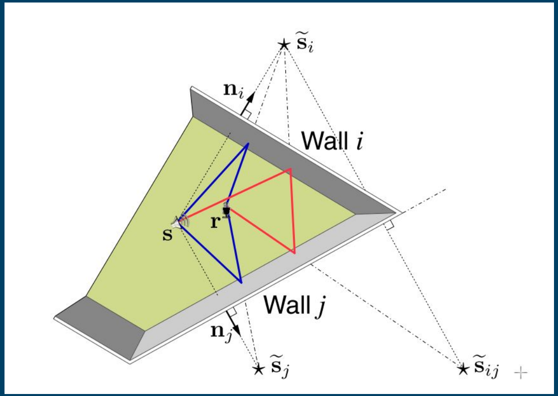


ROOM GEOMETRY

Plan for the Third Year

[Crocco2016,Dockmanic2013, Antonacci2010, Tervo2010]





How to know them?

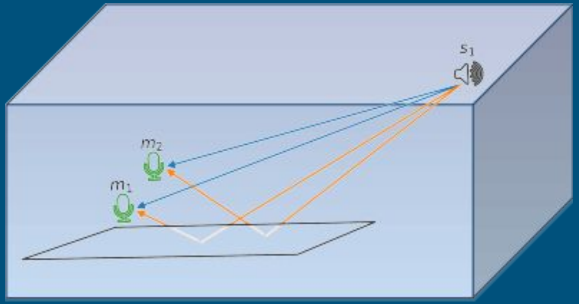
In all the presented works, the echoes are assumed known.

How to estimate them?

Blind Echo Estimation

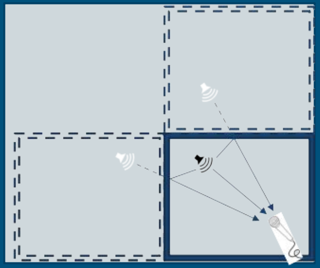
An easy - yet common - scenario: the **pic-nic**

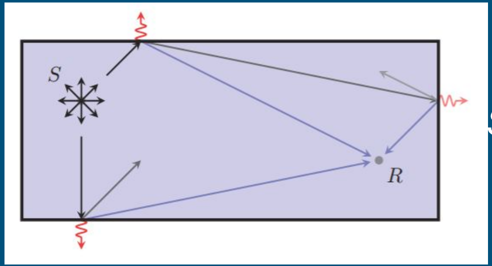
- One Source
- Two microphones
- Random shoe-box rooms
- Nearest surface is the most reflective



10’000 Auditory pic-nic scenes generated using *[Schimmel et al. 2009]* software

Specular reflection





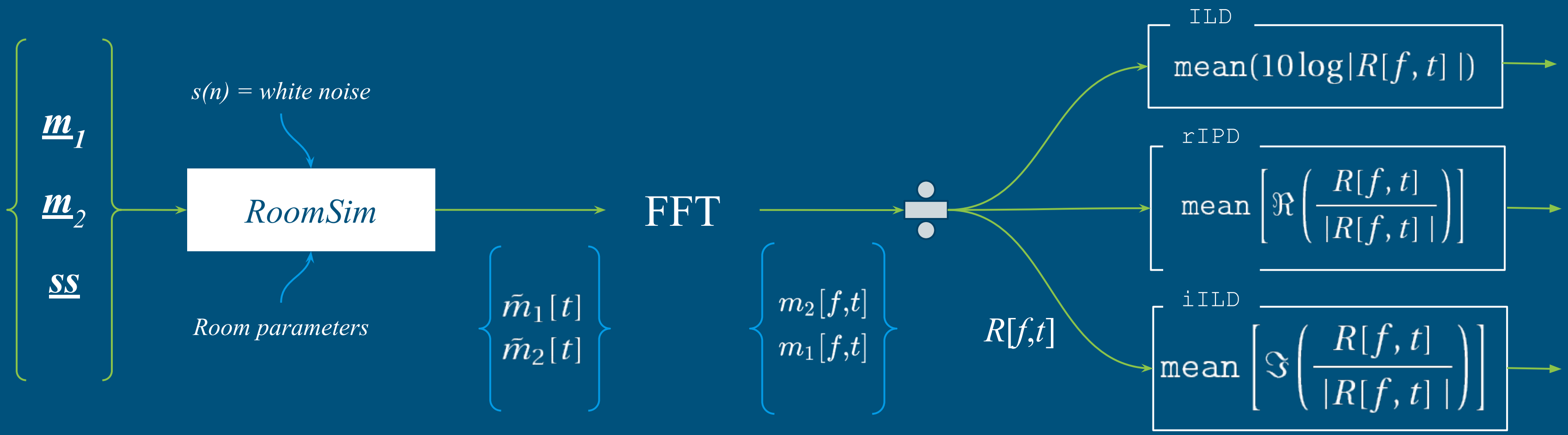
Scattering

[ICASSP19]

Blind Echo Estimation

Why only two microphones?

- The relative transfer function can be computed

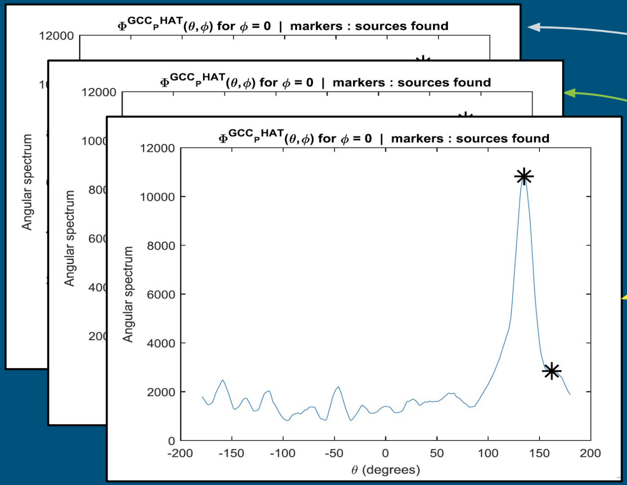


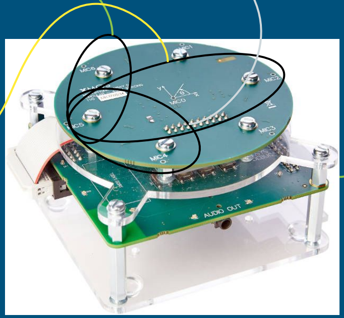
[ICASSP19]

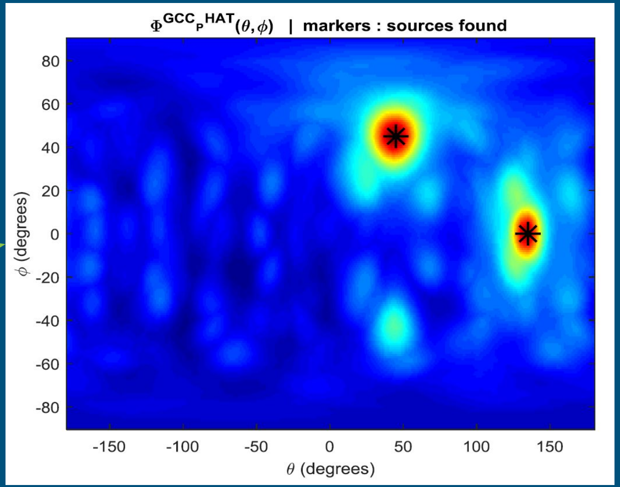
Blind Echo Estimation

Why only two microphones?

- The contribution of multiple microphone pairs can be aggregated together
    - If the geometry of the microphone array is known a priori [*MBSSLocate, DiBiase et al 2001*]







[ICASSP19]

Blind Echo Estimation

Why only two microphones?

- The relative transfer function can be computed

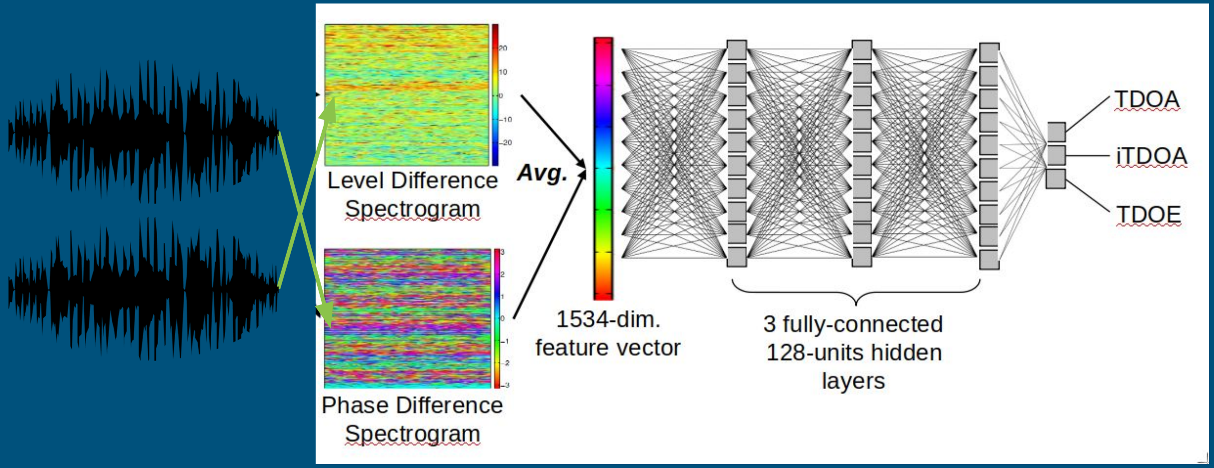
$$\begin{cases} \tilde{m}_1[t] = h_1[t] * \tilde{s}[t] \\ \tilde{m}_2[t] = h_2[t] * \tilde{s}[t] \end{cases} \Rightarrow \quad R[f,t] = \frac{m_2[f,t]}{m_1[f,t]} = \frac{h_2[f]s[f,t]}{h_1[f]s[f,t]} = \frac{h_2[f]}{h_1[f]}$$

  - Ideally it removes the dependency from the source signal
  - If there is no noise **and** filter shorter than the fft window

[ICASSP19]

Blind Echo Estimation - Learning approach

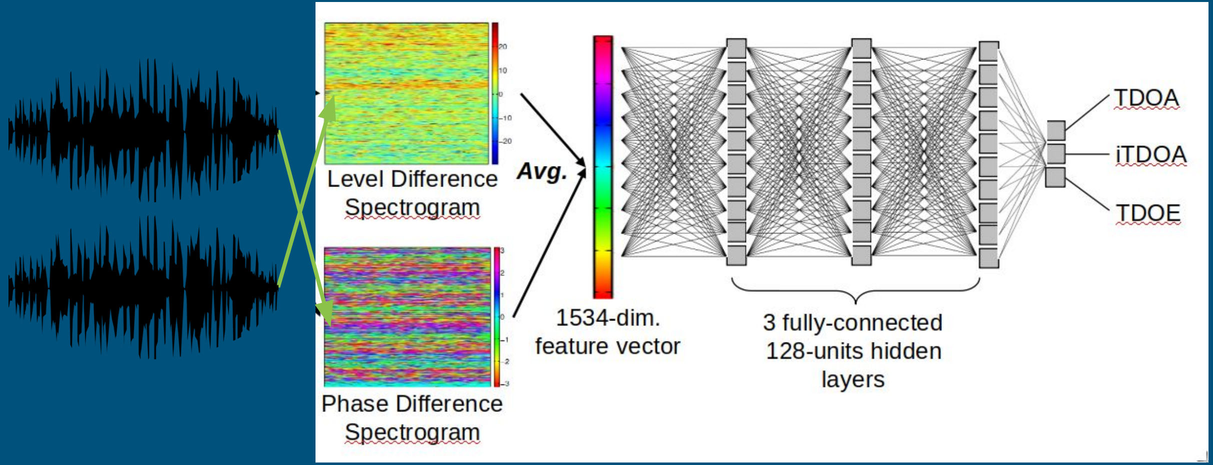
Deep Neural Network Learning



[ICASSP19]

Blind Echo Estimation - Learning approach

Deep Neural Network Learning



Results on test set
[Submitted to ICASSP19]

	Input	nRMSE TDOA	iTDOA	TDOE
MIRAGE	wn	0.18	0.28	0.25
MIRAGE	wn+n	0.68	0.69	0.89
MIRAGE	sp	0.31	0.34	0.56
MIRAGE	sp+n	0.99	0.98	1.48
GCC-PHAT	wn	0.21	-	-
GCC-PHAT	wn+n	0.68	-	-
GCC-PHAT	sp	0.32	-	-
GCC-PHAT	sp+n	1.38	-	-

Also tried with a Gaussian Locally-Linear Mapping (GLLiM) ⇒ It failed

[ICASSP19]

BRAIRE

Blind and constrained acoustic room impulse response estimation

Collaboration with Clement Elvira:

- *An application for the theory of Super Resolution / Continuous Dictionary*
- *Threat as off-grid spike-retrieval*

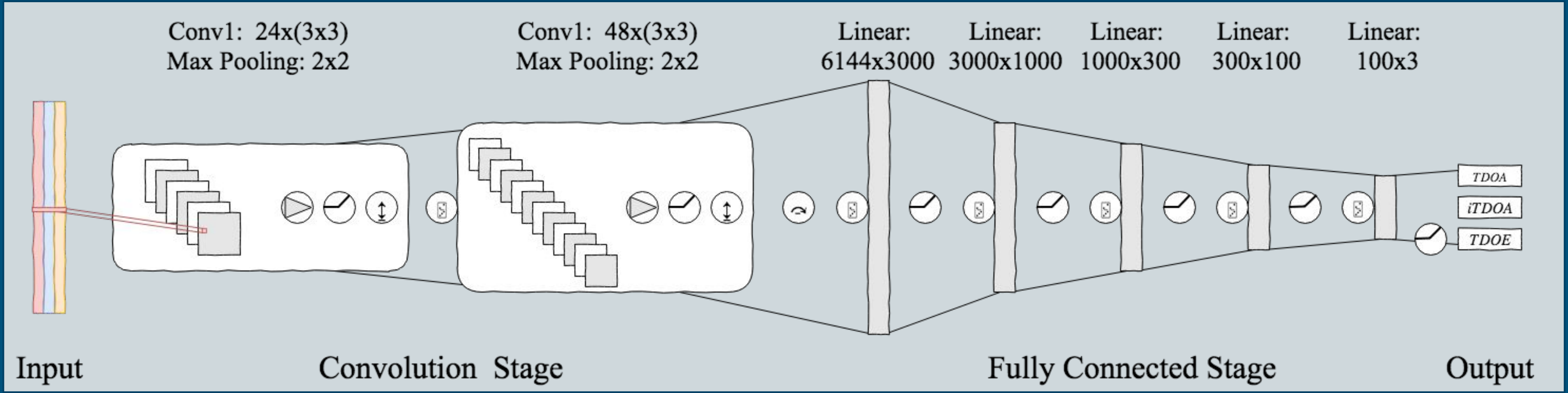
The Hunt Continues...

Current research

1. State of the art DNN architectures
2. *Ad-hoc loss functions for uncertentanetis*
3. Extensions to microphone arrays
4. *Test on real world data*
   a. *Honda Haru*

Echo hunting continues...

- What's next? What's now?
    - State of The Art DNN architecture: CNN [*Chakrabarty et al 2017, Nguyen et al. 2018*]



Echo hunting continues...

- What's next? What's now?
  - State of The Art DNN architecture: CNN
  - Gaussian and Student-T likelihood Loss Function for estimating both TDOA, iTDOA and TDOE and their uncertainties

$$\mathcal{L}(\theta) = \frac{1}{3}\sum_{n=1}^{N} \left|\tau_{a,n} - \tilde{\tau_{a,n}}\right|^2 + \left|\tau_{i,n} - \tilde{\tau_{i,n}}\right|^2 + \left|\tau_{e,n} - \tilde{\tau_{e,n}}\right|^2$$

Echo hunting continues...

- What's next? What's now?
  - State of The Art DNN architecture: CNN
  - Gaussian and Student-T likelihood Loss Function for estimating both TDOA, iTDOA and TDOE and their uncertainties

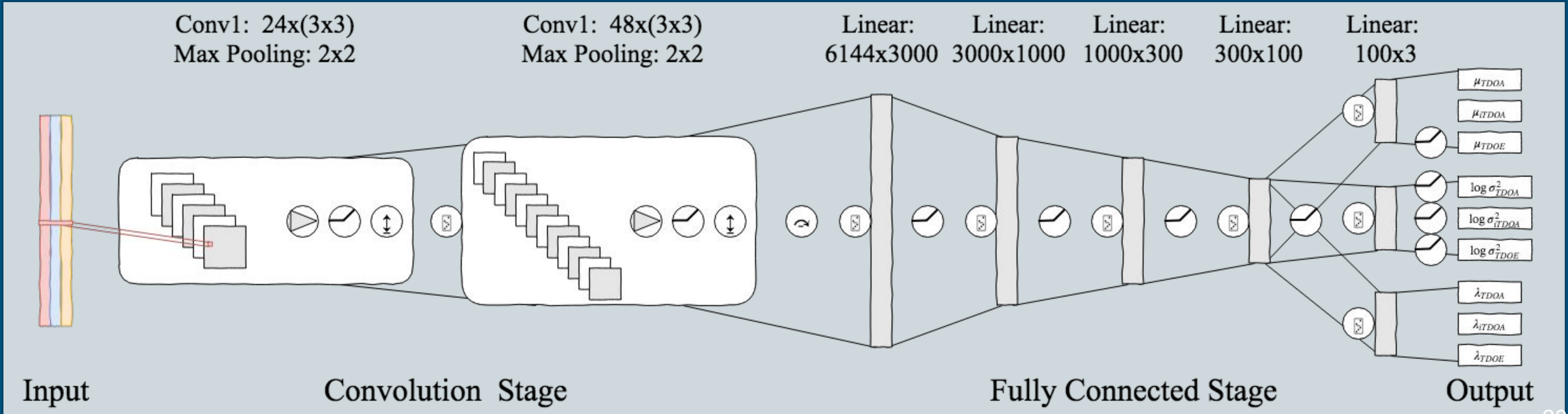
$$\mathcal{L}(\theta) = \tfrac{1}{3}\sum_{n=1}^{N} \left|\tau_{a,n} - \tilde{\tau_{a,n}}\right|^2 + \left|\tau_{i,n} - \tilde{\tau_{i,n}}\right|^2 + \left|\tau_{e,n} - \tilde{\tau_{e,n}}\right|^2$$

$$p(\tau_k | X; \theta) \sim \mathcal{N}(\mu_{\tau_k}(x_n; \theta), \sigma^2_{\tau_k}(x_n; \theta)) \quad k = a, i, e$$

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \log \sigma^2_{\tau_a}(x_n) + \frac{|\tau_a - \mu_{\tau_a}(x_n)|^2}{\sigma^2_{\tau_a}(x_n)} + \ldots$$

Echo hunting continues...

- What's next? What's now?
  - State of The Art DNN architecture: CNN
  - Gaussian and Student-T likelihood Loss Function for estimating both TDOA, iTDOA and TDOE and their uncertainties



Echo hunting continues...

- What's next? What's now?
  - State of The Art DNN architecture: CNN
  - Gaussian and Student-T likelihood Loss Function

Results on test set
[Accepted at ICASSP19]

	Input	nRMSE TDOA	iTDOA	TDOE
MIRAGE	wn	0.18	0.28	0.25
MIRAGE	wn+n	0.68	0.69	0.89
MIRAGE	sp	0.31	0.34	0.56
MIRAGE	sp+n	0.99	0.98	1.48
GCC-PHAT	wn	0.21	-	-
GCC-PHAT	wn+n	0.68	-	-
GCC-PHAT	sp	0.32	-	-
GCC-PHAT	sp+n	1.38	-	-

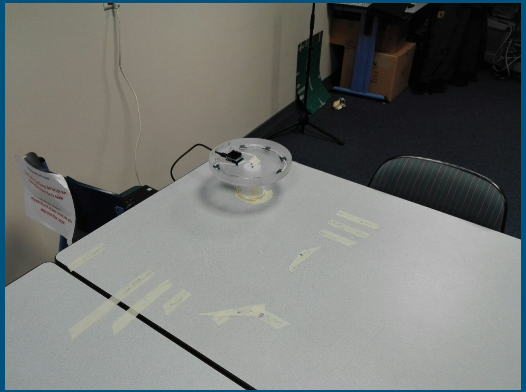
Current Results on test set
with noise

distr	snr	phase	test_signal	tdoa	itdoa	tdoe1
gaussian	0	Test	noise	0.103131	0.110806	0.248462
gaussian	15	Test	noise	0.102640	0.110342	0.280237
gaussian	30	Test	noise	0.101439	0.108265	0.323202
none	0	Test	noise	0.137354	0.145333	0.209920
none	15	Test	noise	0.192951	0.196020	0.284383
none	30	Test	noise	0.148980	0.151179	0.222592
student	0	Test	noise	0.099268	0.107615	0.237591
student	15	Test	noise	0.110567	0.111748	0.310297
student	30	Test	noise	0.106170	0.113793	0.294742

Echo hunting continues...

- What’s next? What’s now?
  - State of The Art DNN architecture: CNN
  - Gaussian and Student-T likelihood Loss Function
  - Test on real data and microphone array (HONDA HARU array)





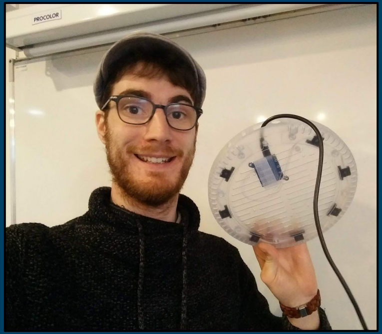
HONDA HARU

An application for MIRAGE

Submitted to HONDA on March 2019

Phase II: passed
Phase III: next year

Research project funded by HONDA



SPCUP2019: DREGON

Drone ego noise reduction for sound source localization

[ICASSP 2019]

IEEE Signal Processing Cup 2019

Search and Rescue via Drone-Embedded Sound Source Localization

Student Competition

Organizing committee:




MIRA for DOLBY

Music Interference Reduction Algorithm

Submitted to DOLBY on December 2018
Project Accepted

Algorithm for microphone leakage removal on full-length audio recordings

No Echo model at All : Phase is considered Random

Tasks:

- Soundcheck dataset
  - Learning with sound-check
  - Removal on the actual song
- Gorillaz concert
  - Remove PA from Audience
- Football match
  - Commentator enhancement
  - Ball enhancement

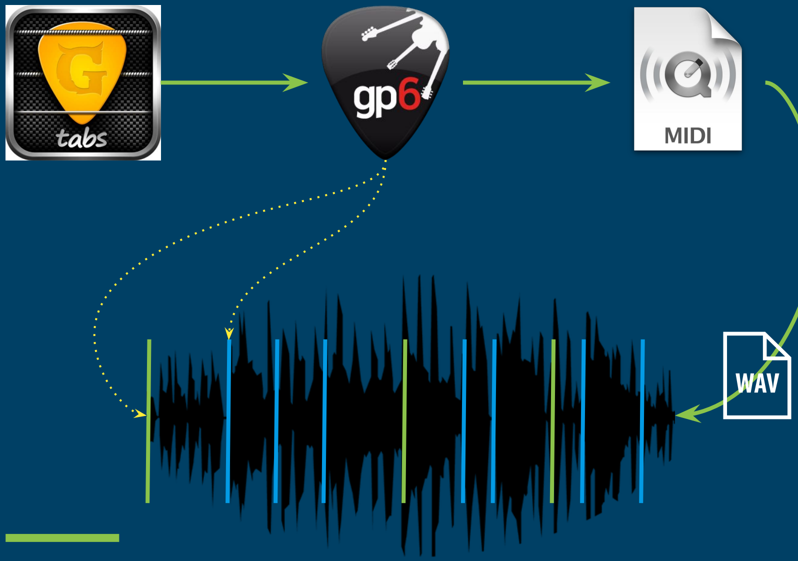
BEATLESS

downbeat Detection by Learning on Synthetic Sources

Collaboration with Magdalena Fuentes
Telecom ParisTech/CentraleSupelec

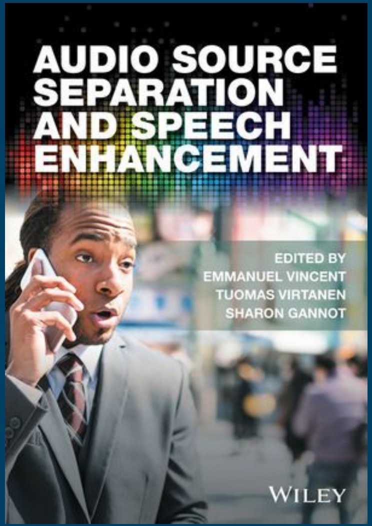
Virtually-Supervised
Learning-based

Downbeat detection and micro-timing



Prof. GANNOT

Visiting Bar'Ilan University
November 2019 - February 2020



RAPPLE

Rap APP battLE

THANK YOU

Need for some answers?