Immersive Audiovisual Production Enhancement based on 3D Audio

PhD Research Proposal

Andrés Pérez López

September 8, 2017

Outline

Outline for Section 1

Context

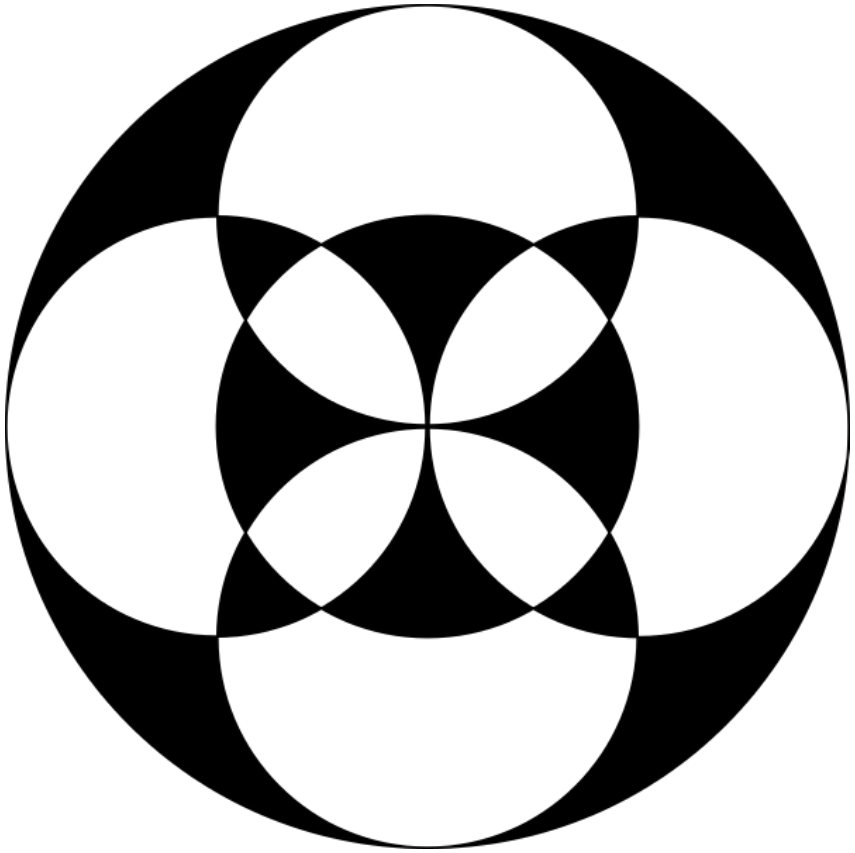
Eurecat

upf. Music Technology Group

eurecat Centre Tecnològic de Catalunya

Motivation

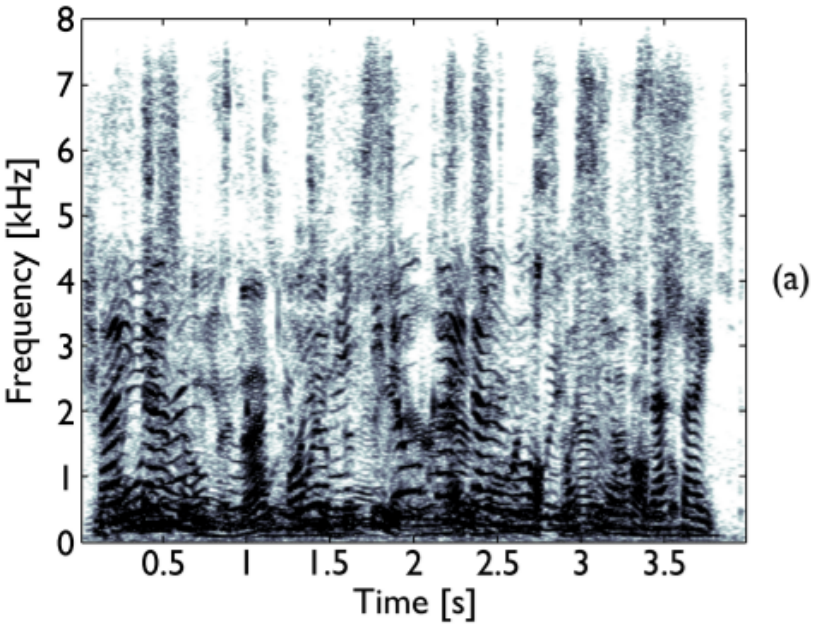
Ambisonics

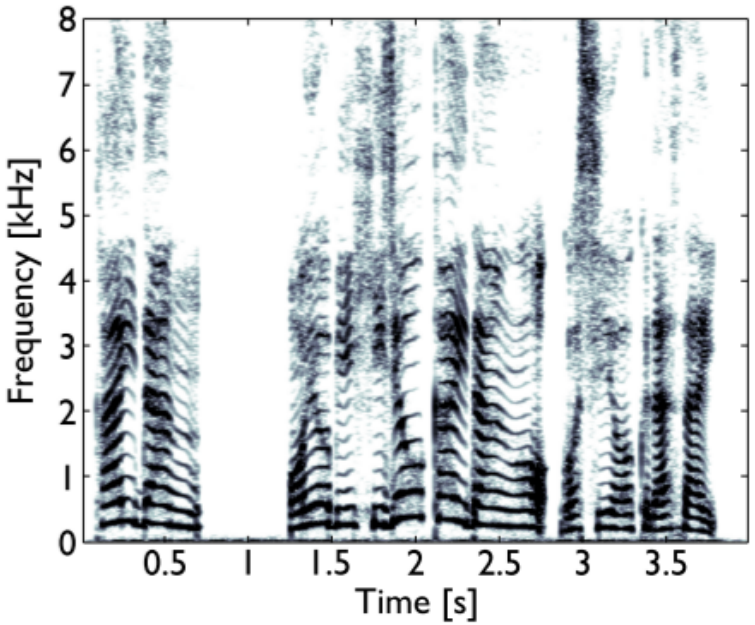


AnonMoos, *Wikipedia: Ambisonics*

Motivation

Blind Source Separation





Adapted from N. Epain *et. al.*, *Blind Source Separation using Independent Componen Analysis in the Spherical Harmonics Domain*, 2010

Motivation

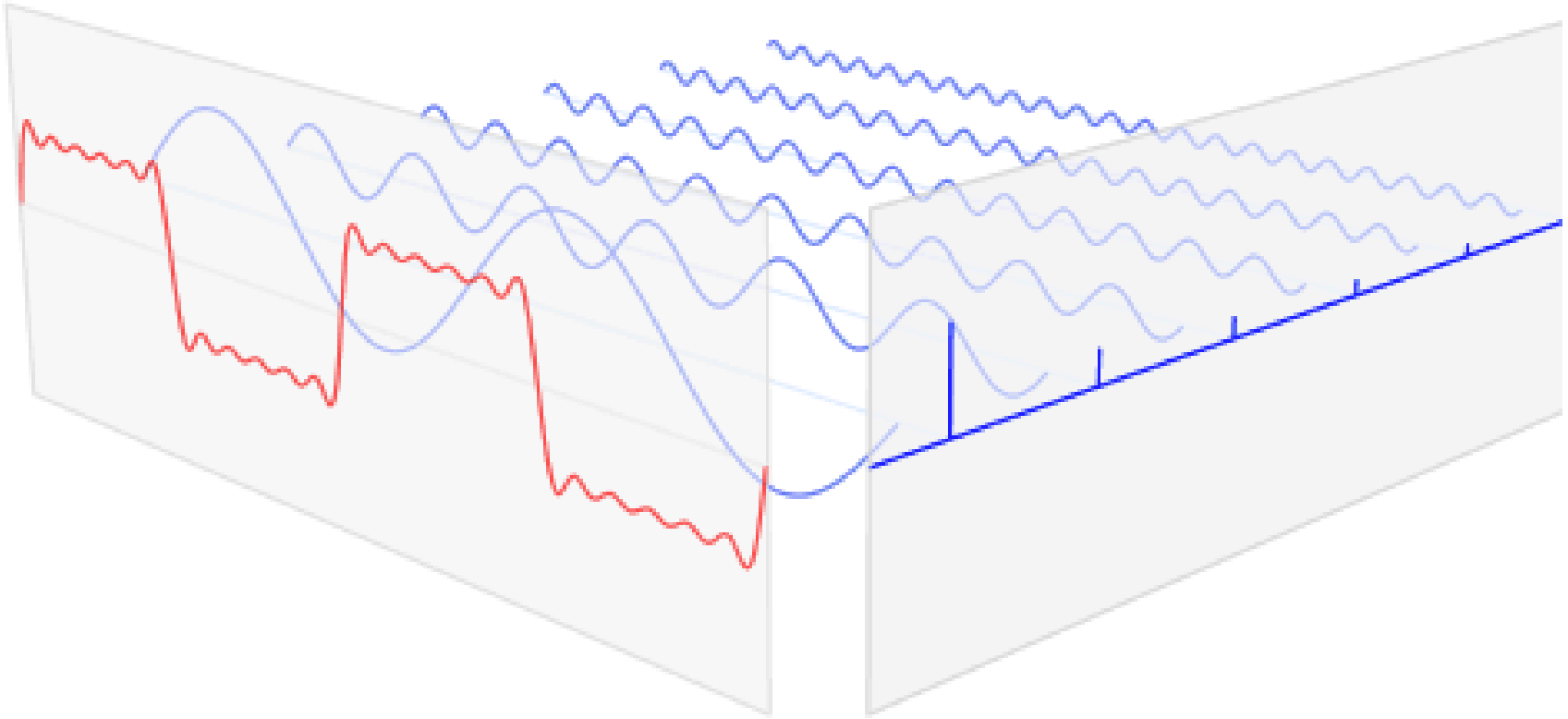
Idea

Multichannel spatial information contained in Ambisonics audio might be exploited by Blind Source Separation algorithms.

Outline for Section 1

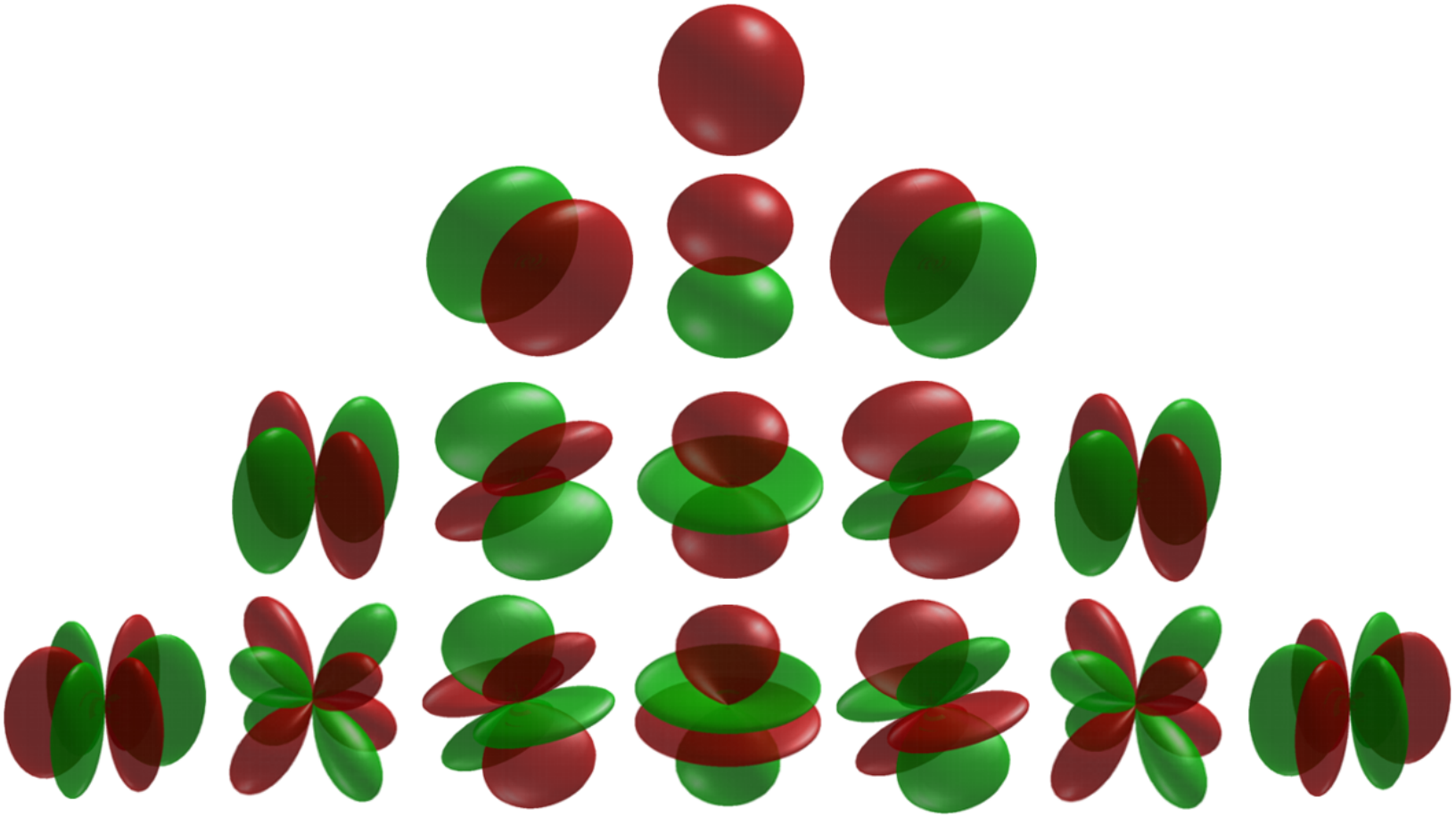
Ambisonics

Theory



Ambisonics

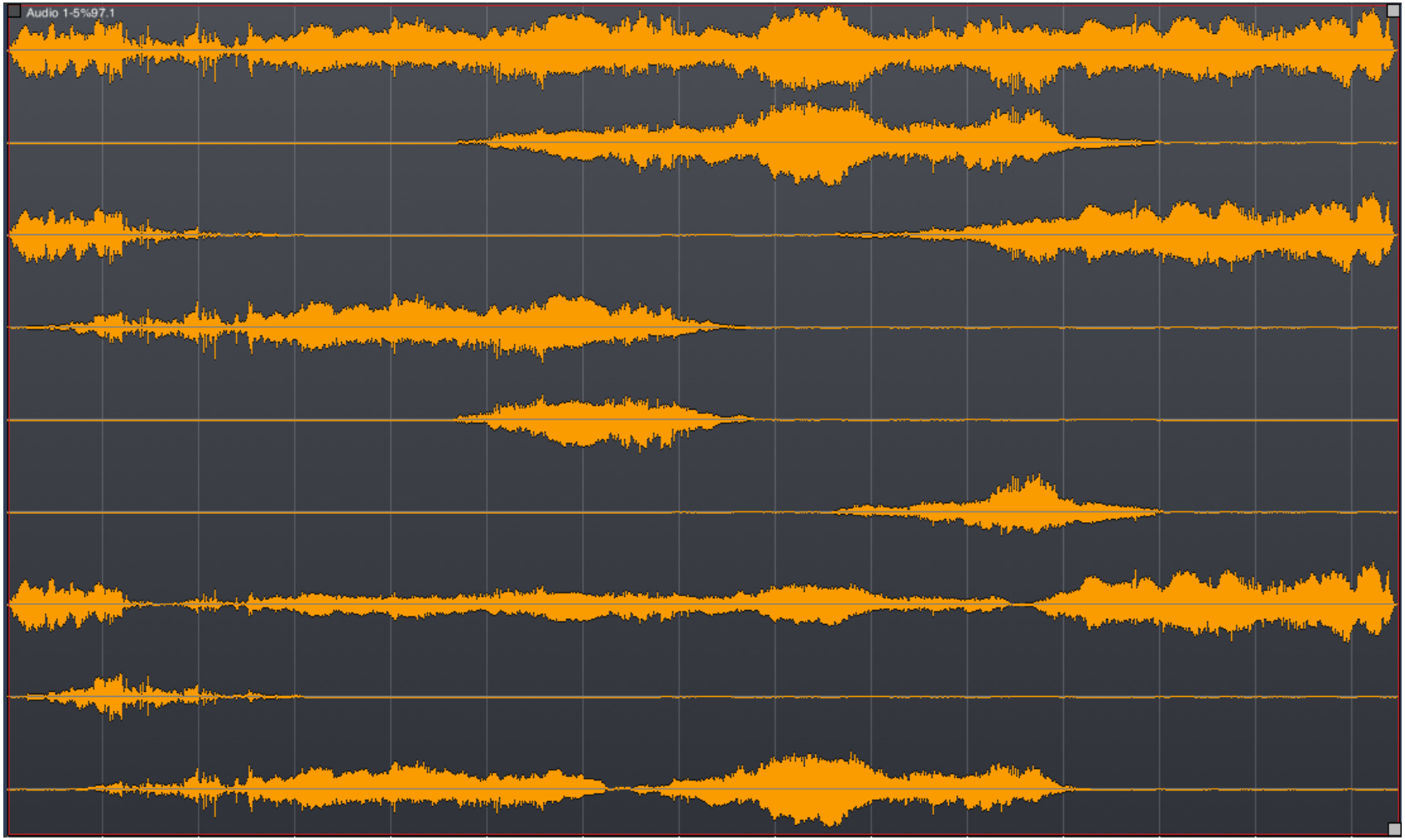
Theory



Sarxos, *Wikipedia: Spherical Harmonics*

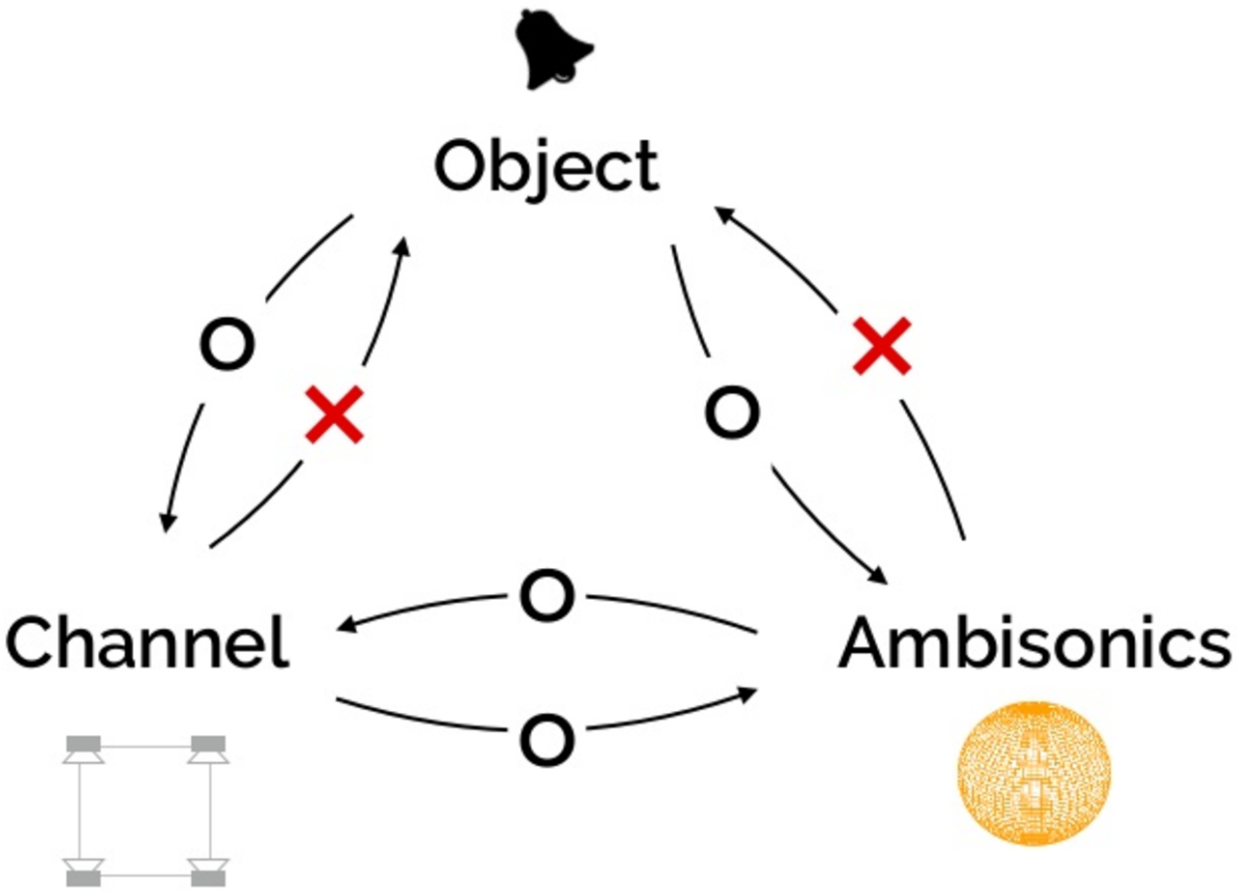
Ambisonics

Theory



Ambisonics

Formats



Adapted from GAudio , *End-to-End VR Audio Solution for Fully Immersive/Interactive Experience*, 2016

Ambisonics

HRTF - Binaural

NoiseMakers, *Ambi Head*

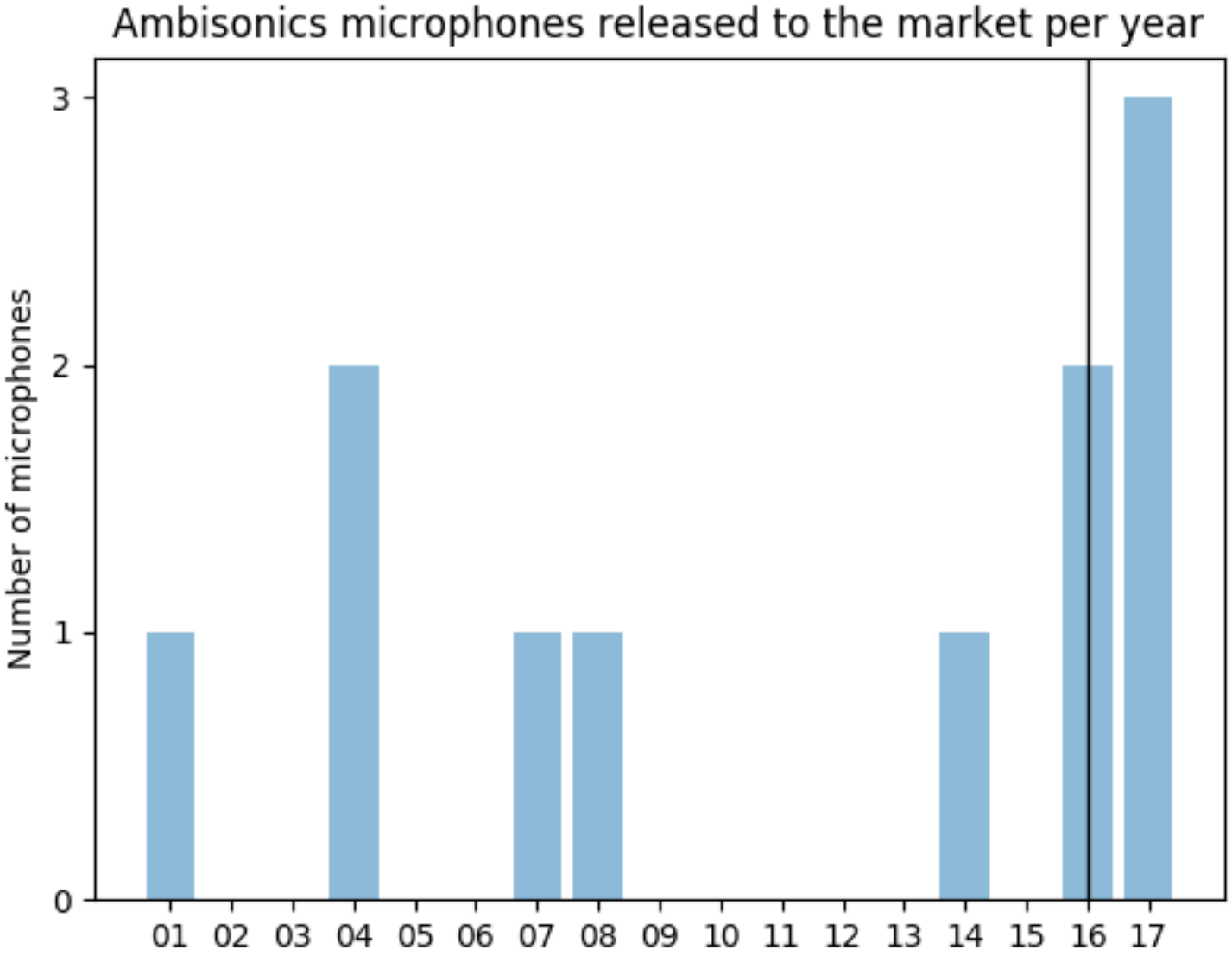
Ambisonics

Recording

From E. Bates *et. al.*, *Comparing Ambisonics Microphones - Part 2*, 2017

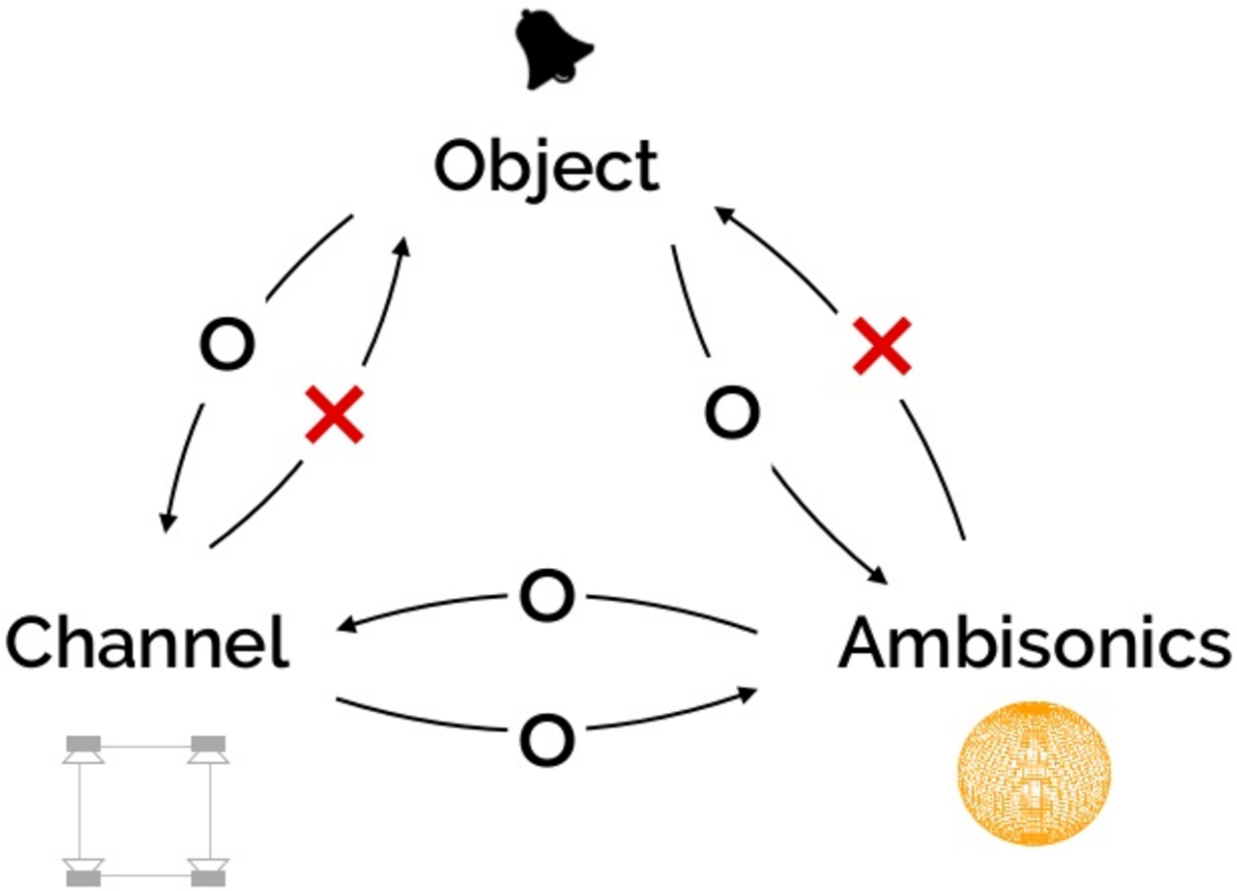
Ambisonics

Why?



Ambisonics

Why?

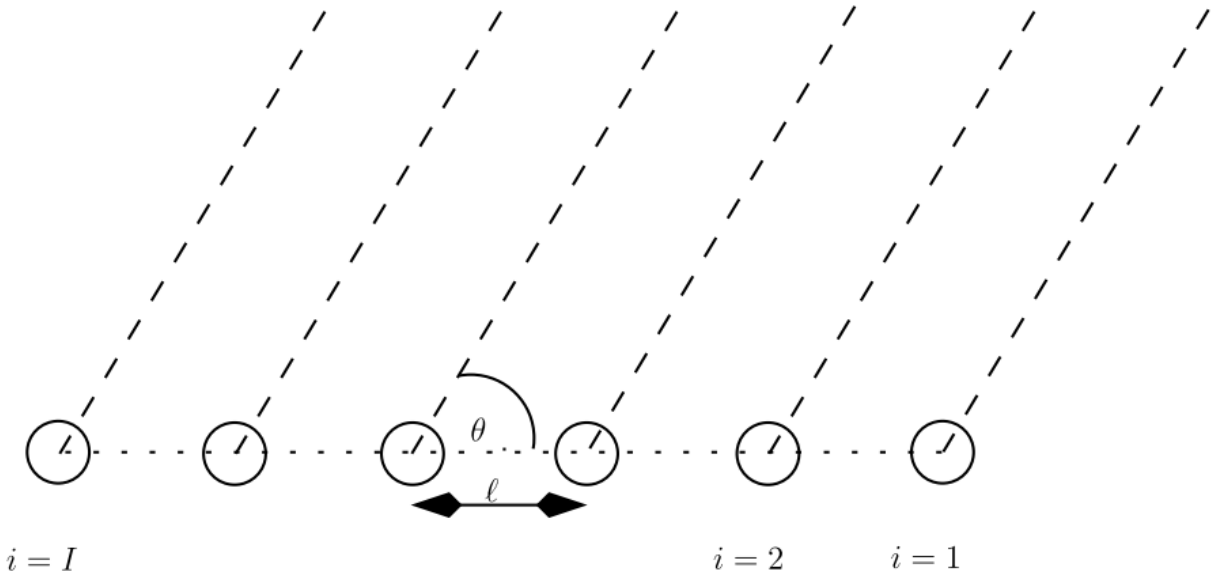


Adapted from GAudio , *End-to-End VR Audio Solution for Fully Immersive/Interactive Experience*, 2016

Outline for Section 2

Sound Source Localization

Linear Arrays - TDoA



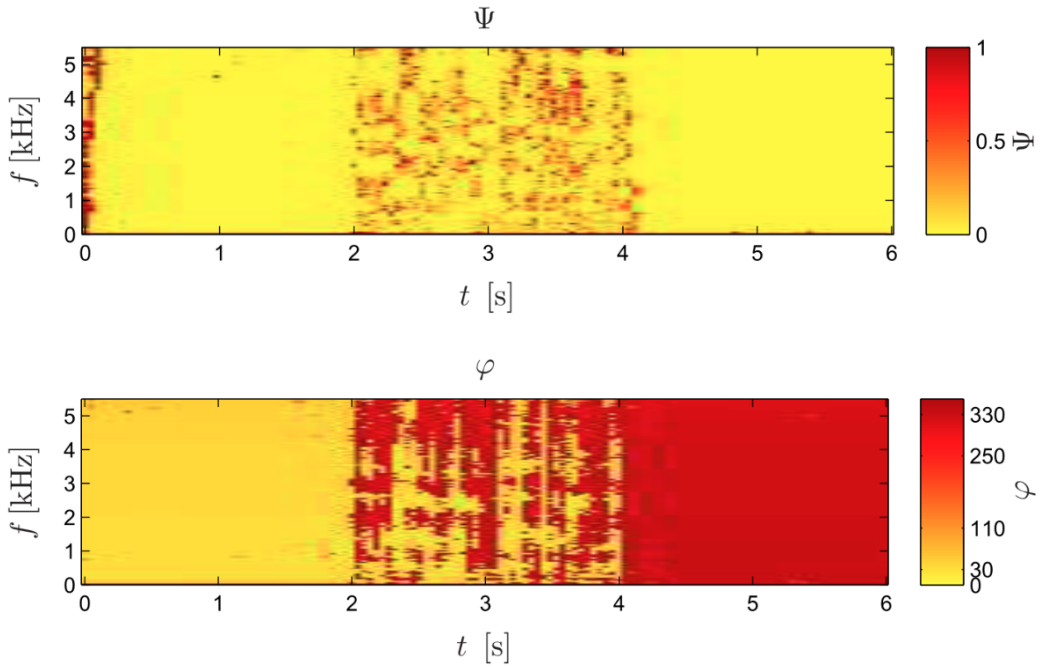
From A. Gannot *et. al.*, *A consolidated perspective on multi-microphone speech enhancement and source separation*, 2017

Sound Source Localization

Ambisonics Intensity Vector Analysis

$$\Psi(k,n) = 1 - \frac{\| < \boldsymbol{I}_{\mathrm{a}}(k,n) >_t \|}{< \|\boldsymbol{I}_{\mathrm{a}}(k,n)\| >_t}$$

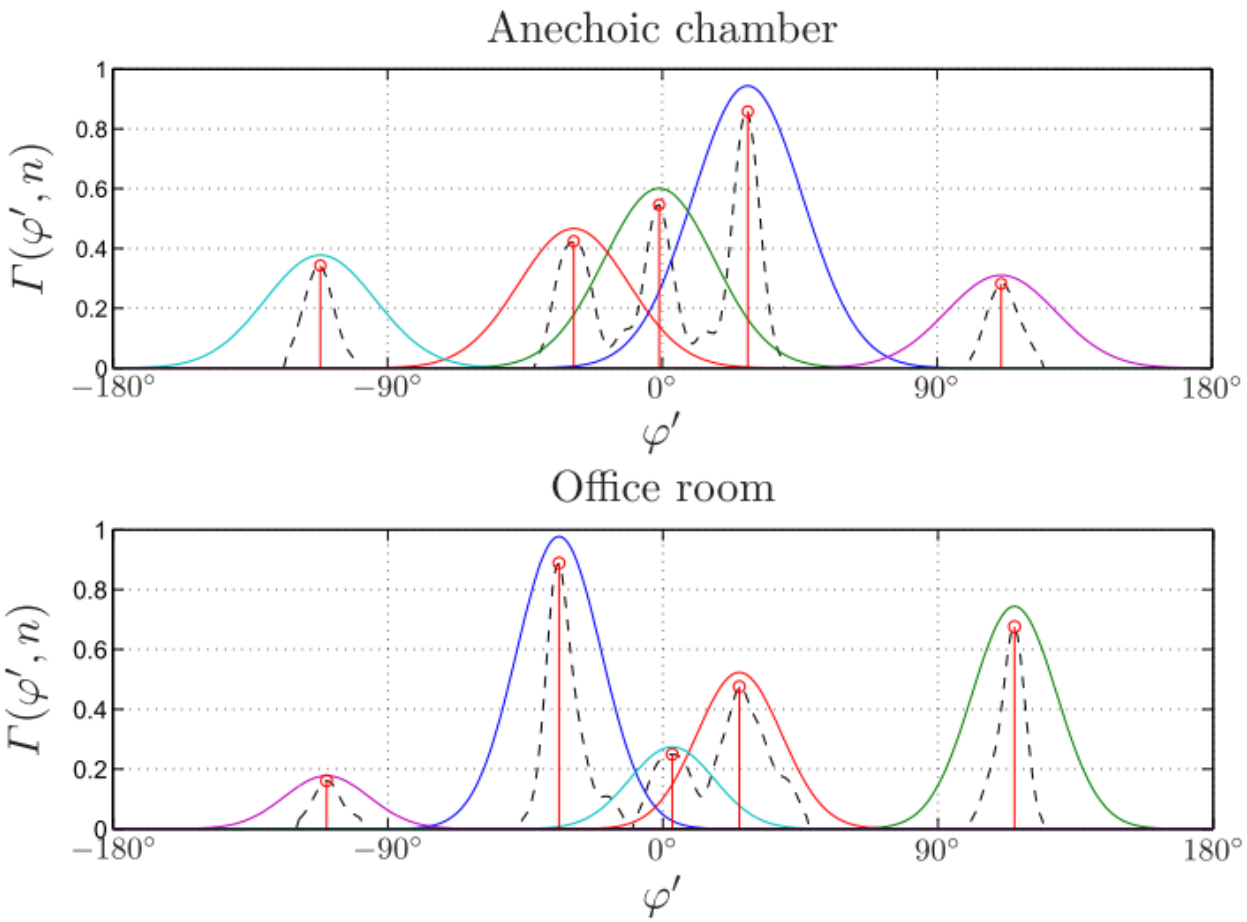
$$\boldsymbol{I}_{\mathrm{a}}(k,n) = \frac{1}{2}\,\Re\left\{P(k,n)\cdot\overline{\boldsymbol{U}(k,n)}\right\}$$



Adapted from O. Thiergart *et. al.*, *Localization of Sound Sources in Reverberant Environments Based on Directional Audio Coding Parameters*, 2009

Sound Source Localization

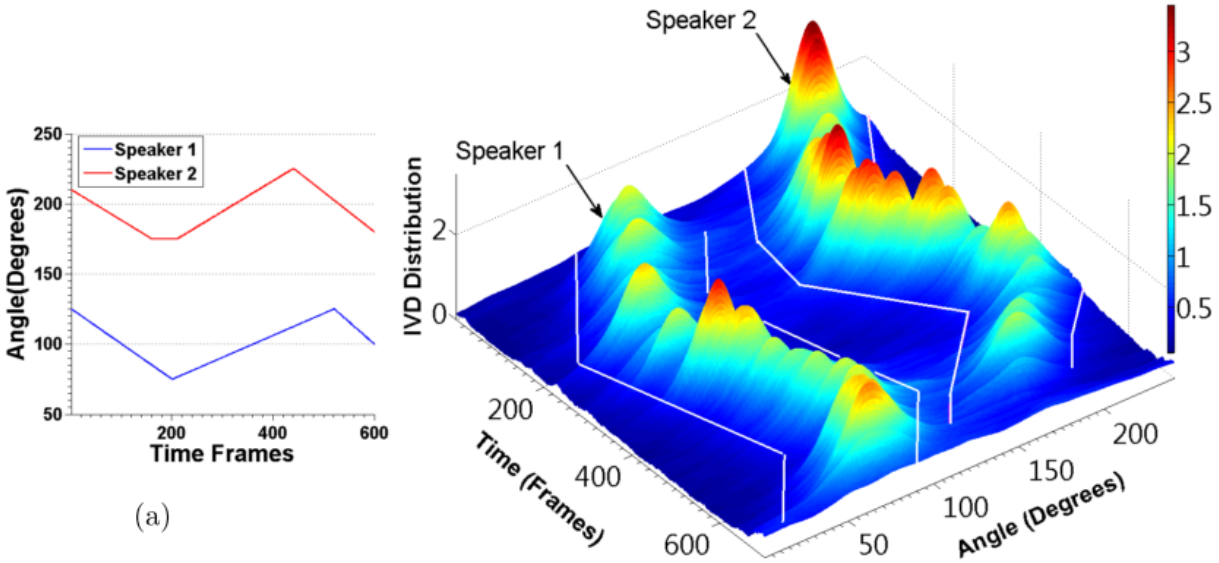
B-Format Intensity Vector Analysis



From O. Thiergart *et. al.*, *Localization of Sound Sources in Reverberant Environments Based on Directional Audio Coding Parameters*, 2009

Sound Source Localization

B-Format Intensity Vector Analysis



From A. Riaz, *Adaptive Blind Source Separation Based on Intensity Vector Statistics*, 2015

Sound Source Localization

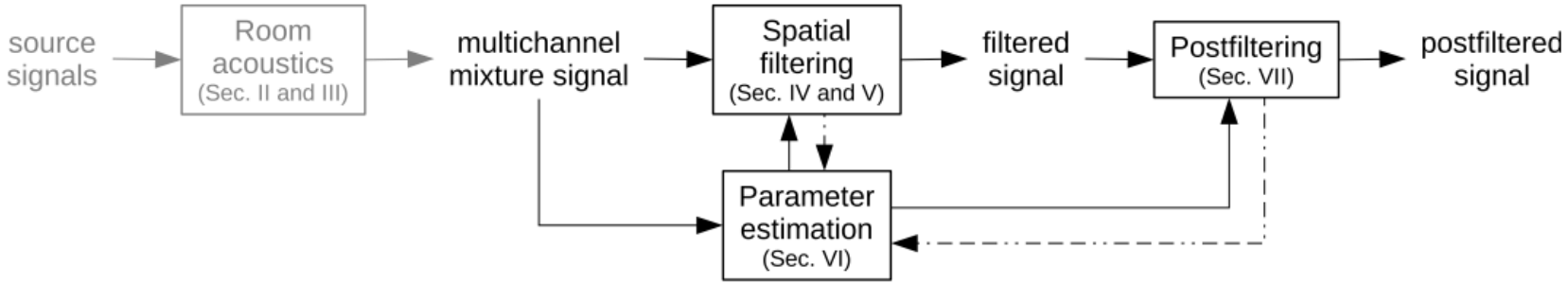
Ambisonics-SSL review

Article	Method	Ambisonics Order	Microphone	Number of Capsules
Pulkki07 [19]	IV	1	-	-
Thiergart09 [21]	IV + GMM	1 horizontal	Custom circular	4
Tervo09 [22]	IV + vonMises MM	1 horizontal	Custom circular	4
Pavlidi15 [23]	IV + SSZ	1	Custom spherical	32
Pulkki13 [24]	Sectorial IV	HOA	-	-
He17 [25]	IV + local DOA + accuracy + FOSDA	1 horizontal	Custom circular	4
Ding17 [26]	IV + local DOA + accuracy + KMeans	1 horizontal	Custom circular	4
Jarret10 [27]	PIV	HOA	Eigenmike	32
Evers14 [28]	PIV + K-Means	HOA	Custom spherical	32
Moore15 [29]	PIV + DPD	HOA	Custom spherical	32
Nadiri14 [30]	PWD + SCM + DPD	HOA	Eigenmike	32
Berge10 [31]	Harpex	1	-	-
Thiergart12 [32]	Harpex	1	-	-
Dimoulas07 [34]	A-EBL	1	SoundField	4
Dimoulas09 [35]	(DWT/SWT)-JTF-A-EBL	1	SoundField	4

Outline for Section 2

Blind Source Separation

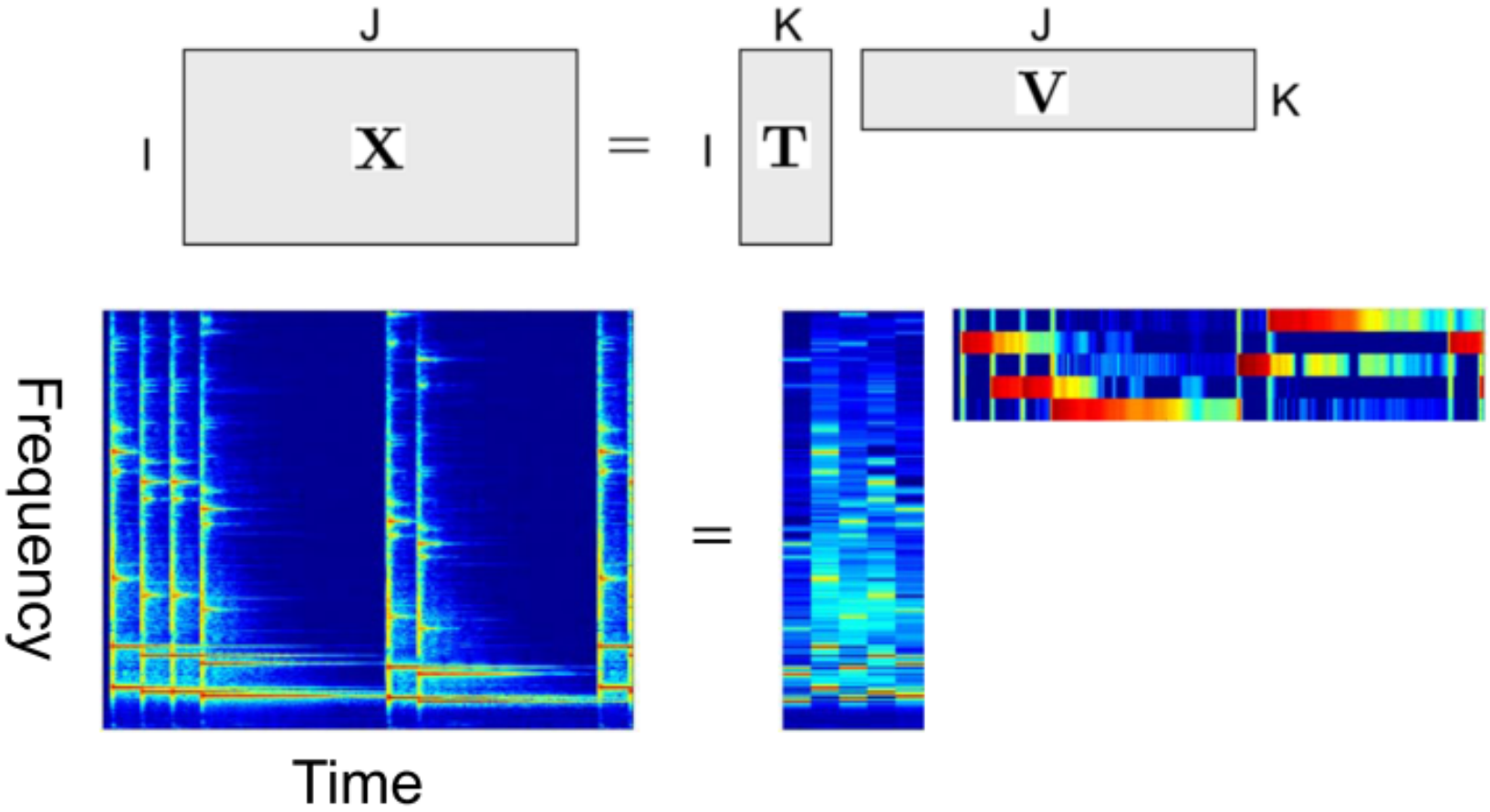
Multichannel



From S. Gannot *et. al.*, *A Consolidated Perspective on Multi-Microphone Speech Enhancement and Source Separation*, 2017

Blind Source Separation

Multichannel



From H. Sawada *et. al.*, *Multichannel Extensions of Non-negative Matrix Factorization with Complex-valued Data*, 2013

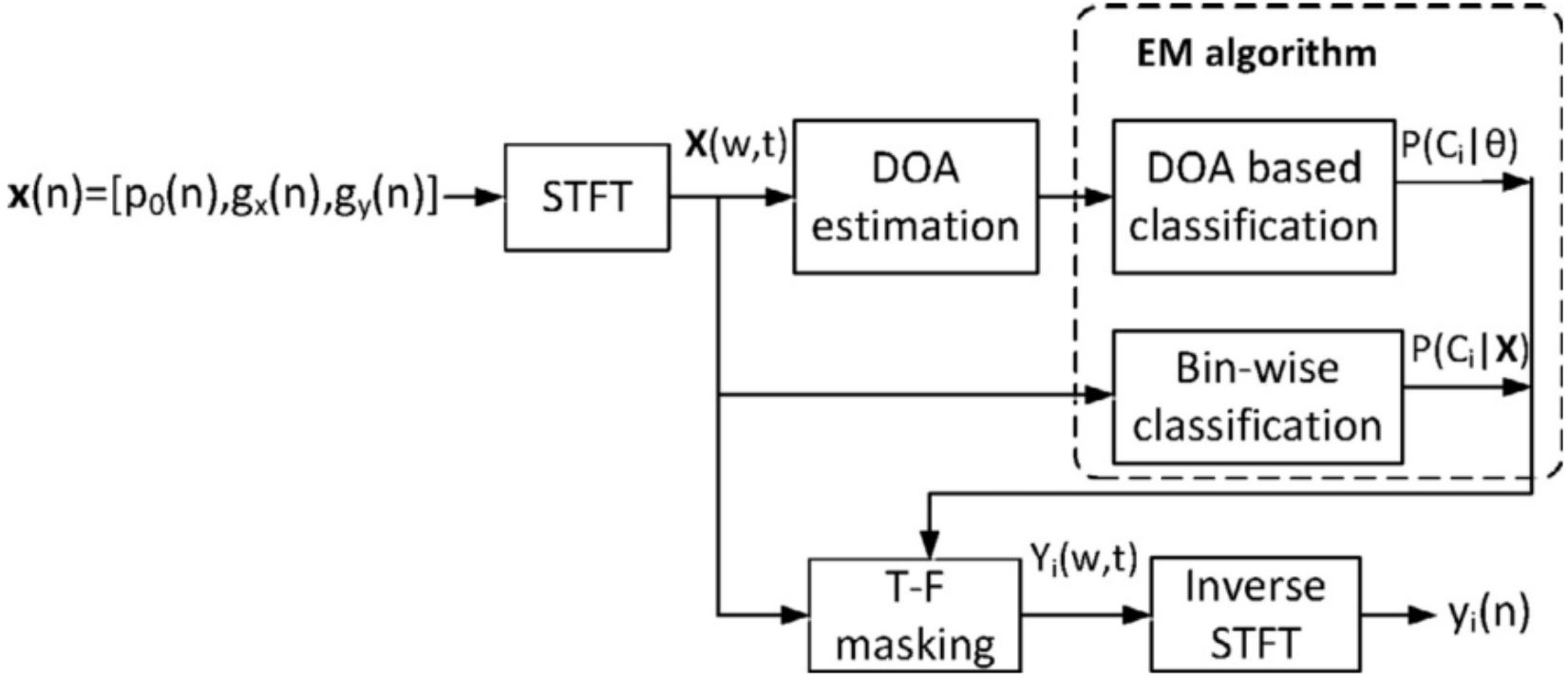
Blind Source Separation

Multichannel BSS review

Article	Method	Microphone Array	L	Target Sound	Dataset	Evaluation Metrics
Epain10 [40]	ICA	Custom spherical	2	Speech	custom	PESQ
Baque16 [41]	ICA (ERBM)	Custom spherical	2	Speech	custom	SDR, DOA
Ozerov09 [42]	NMF	Linear array	-	Music	SiSEC08	SDR, ISR, SIR, SAR
Duong11 [44]	Gaussian SCM + ML	Linear array	-	Speech	custom	SDR, ISR, SIR, SAR
Arberet11 [46]	Gaussian SCM + NMF	Linear array	-	Music	custom	SDR
Sawada13 [38]	Spatial CNMF	Linear array	-	Music	SiSEC11	SDR

Blind Source Separation

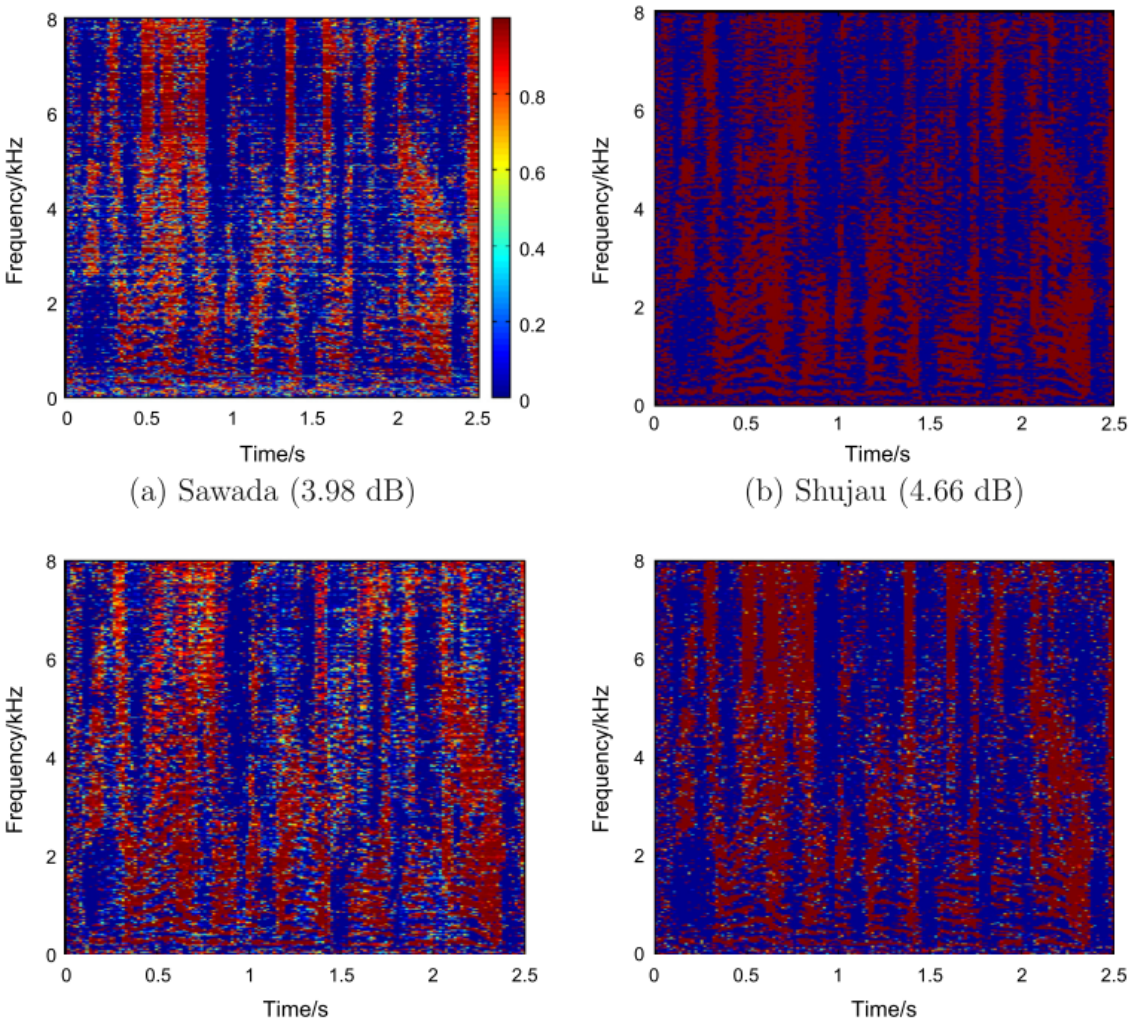
Ambisonics SSL-BSS



From X. Chen *et. al.*, *Reverberant speech separation with probabilistic time-frequency masking for B-format recordings*, 2015

Blind Source Separation

Ambisonics SSL-BSS



(a) Sawada (3.98 dB)

(b) Shujau (4.66 dB)

(c) Gunel (5.62 dB)

(d) Proposed with reliability (6.53 dB)

From X. Chen *et. al.*, *Reverberant speech separation with probabilistic time-frequency masking for B-format recordings*, 2015

Blind Source Separation

Ambisonics SSL-BSS review

Article	Method	Ambisonics Microphone	L	Target Sound	Dataset	Evaluation Metrics
Gunel08 [49]	IV + vonMises MM + Softmask	SoundField	1 h	Speech	Music for Archimedes	SDR, SIR
Riaz15 [50]	Gunel + Mic Correction + Adaptive Filter + Location Estimation	SoundField	1 h	Speech, Music	Music for Archimedes	SDR, ISR, SIR, SAR
Shujau11 [51]	IV + VAD + DOA Clustering + Binary Mask	AVS	1 h	Speech	TIMIT	SDR, ISR, PESQ-MOS
Chen15 [53]	IV + MV + Softmask	SoundField	1 h	Speech	TIMIT	SDR, ISR, PESQ-MOS

Outline for Section 2

Multimodal Enhancement for BSS

Audiovisual SSL

(a) green: audio, blue: visual (b) correct detection (c) observations (d) bad detection

From S. Gebru *et. al.*, *Audio-Visual Speaker Localization via Weighted Clustering*, 2014

Multimodal Enhancement for BSS

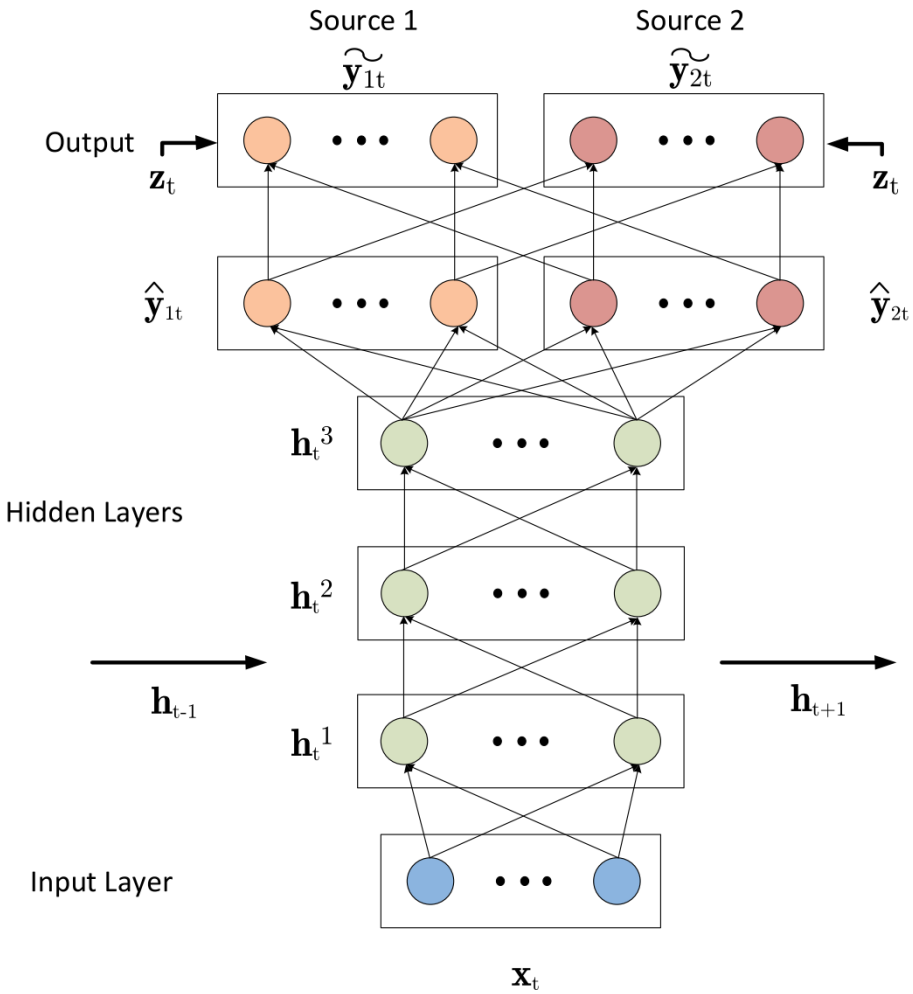
Multimodal SSL/BSS review

Article	Localization Method	Separation Method	Target	Mic	Camera
Khalidov11 [55]	Conjugate GMM	-	Speech	2 omni	2
Gebru14 [56]	Weighted-Data GMM	-	Speech	Binaural head	1
Khan13 [57]	MCMC-PF	GMM-EM	Speech	Binaural head	2

Outline for Section 2

Machine Learning for BSS

Monophonic Musical BSS



From P. Huang *et. al.*, *Singing-Voice Separation from Monaural Recordings using Deep Recurrent Neural Networks*, 2014

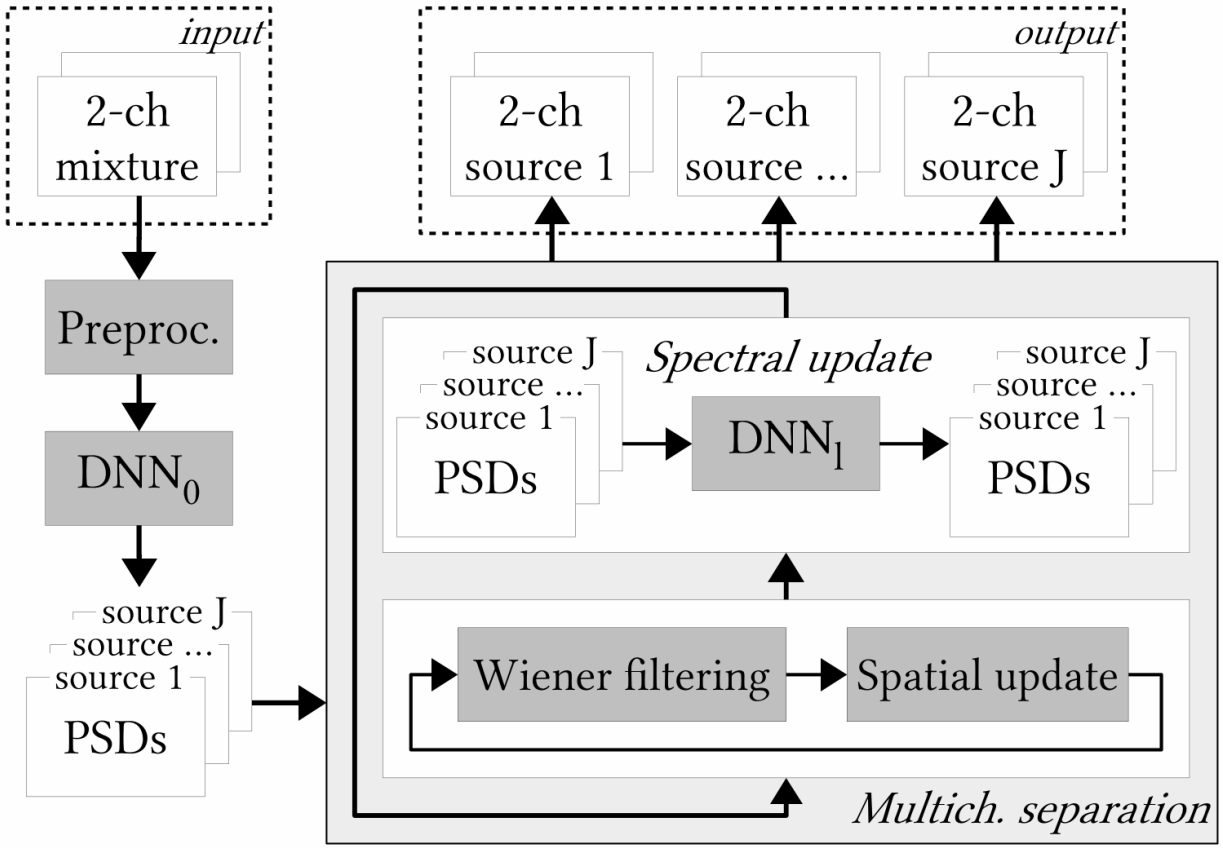
Machine Learning for BSS

Monophonic Musical DNN for BSS review

Article	DNN Architecture	Target	Dataset	Evaluation Metrics
Huang14 [58]	DNN, DRNN, sRNN	Singing Voice	MIR-1k	SDR, SIR, SAR
Uhlich15 [59]	ReLU DNNU	Predefined Instrument	TRIOS	SDR, SIR, SAR
Uhlich17 [60]	Feed-Forward, Bi-LSTM	Vocal, Bass, Drum, other	DSD100	SDR,R
Sebastian16 [61]	MOD-GD DRNN	Singing Voice, Vocal-Violin	MIR-1k	SDR, SIR, SAR
Chandna17 [62]	DNN, 2 convolutional layers	Vocal, Bass, Drum, other	DSD100, MSD100	SDR, SIR, SAR, ISR

Machine Learning for BSS

Multichannel BSS



From A. Nugraha *et. al.*, *Multichannel music separation with deep neural networks*, 2016

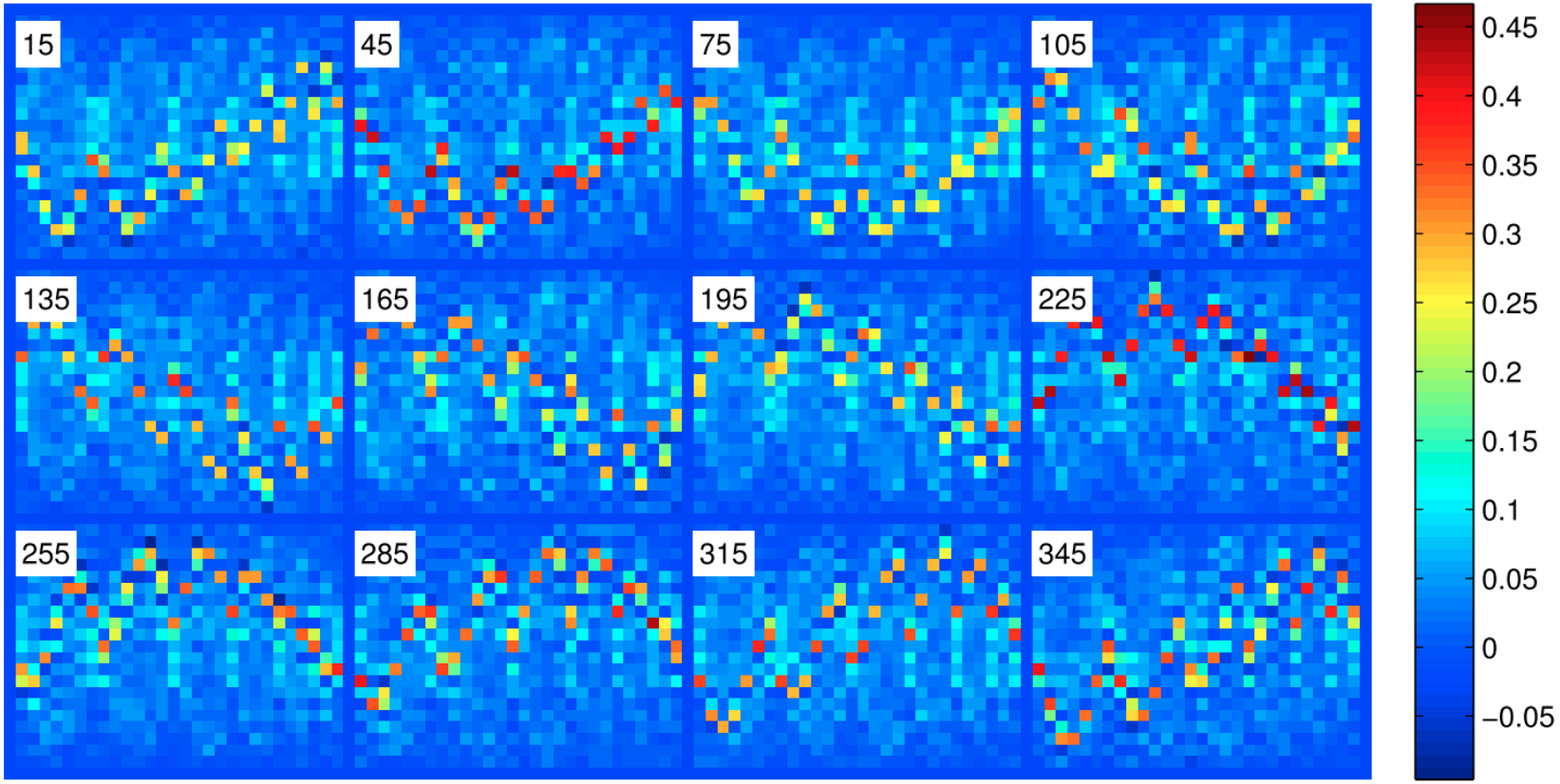
Machine Learning for BSS

Multichannel DNN for BSS review

Article	Method	Microphone Array	#Mics	Target Sound	Dataset	Evaluation Metrics
Nugraha16 [63]	DNN-PSD + SCM	Custom linear	2	Vocals	DSD100	SDR, SIR, ISR, SAR
Nugraha16(2) [64]	DNN-PSD + SCM	Custom linear	6	Speech	CHiME-3	SDR, SIR, ISR, SAR
Wisdom16 [65]	DMCGMM	Custom circular	8	Speech	WSJCAM0 REVERB	SDR
Erruz17 [66]	ILD-CNN	-	2	Music	DSD100	SDR, SIR, ISR, SAR

Machine Learning for BSS

Sound Source Localization



From X. Xiao *et. al.*, *A learning-based approach to direction of arrival estimation in noisy and reverberant environments*, 2015

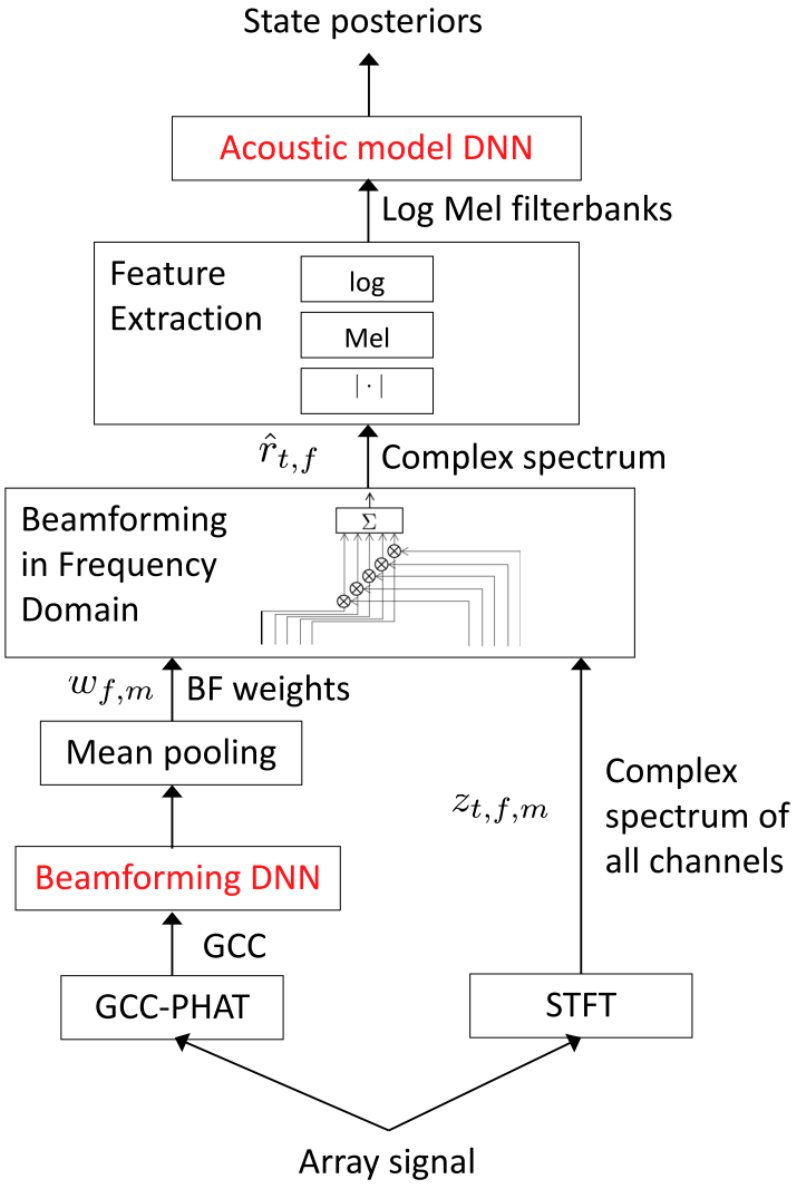
Machine Learning for BSS

Multichannel DNN for SSL review

Article	Method	Microphone Array	#Mics	Target Sound	Dataset	Evaluation Metrics
Xiao15 [67]	MLP GCC-PHAT	Custom circular	8	Speech	WSJCAM0	DOA RMSE, MAE
Chakrabarty17 [68]	phase spectrogram CNN	Custom linear	4	Speech	Synthesized Noise	SDR

Machine Learning for BSS

Multichannel SSL-BSS



From X. Xiao *et. al.*, *Deep Beamforming Networks for Multi-Channel Speech Recognition*, 2016

Machine Learning for BSS

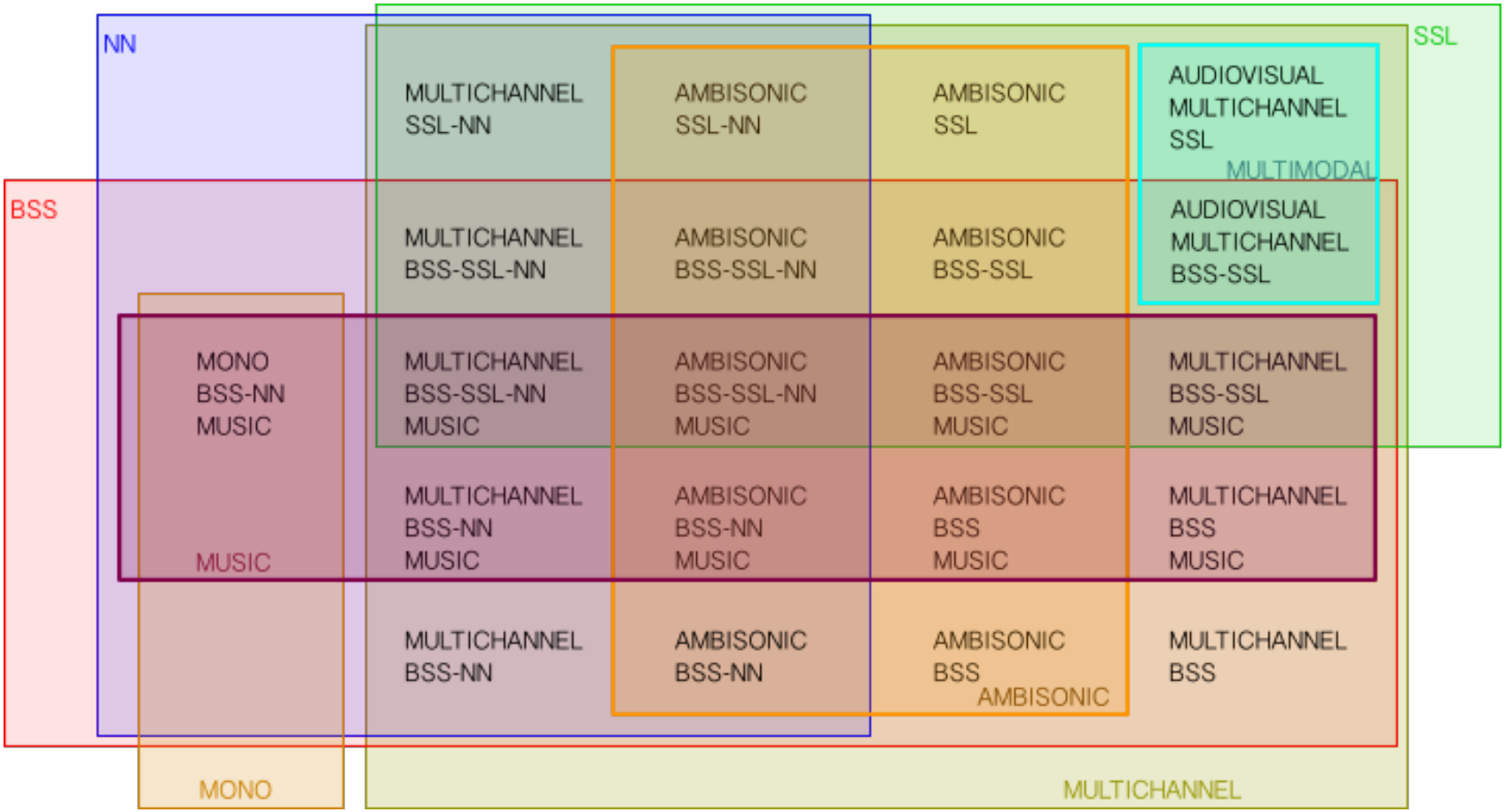
Multichannel DNN for SSL-BSS review

Article	Method	Microphone Array	#Mics	Target Sound	Dataset	Evaluation Metrics
Araki15 [69]	ITD, ILD DAE	Binaural	2	Speech	PASCAL CHiME	SSNR, CD
Jiang14 [70]	ITD, ILD, GFCC DNN	Binaural	2	Speech	Custom speech, ROOM-SIM	HIT, FA, HIT-FA, SNR
Xiao16 [71]	FF GCC-PHAT, LSTM AM	Custom circular	8	Speech	WSJCAM0, REVERB	WER

Outline for Section 2

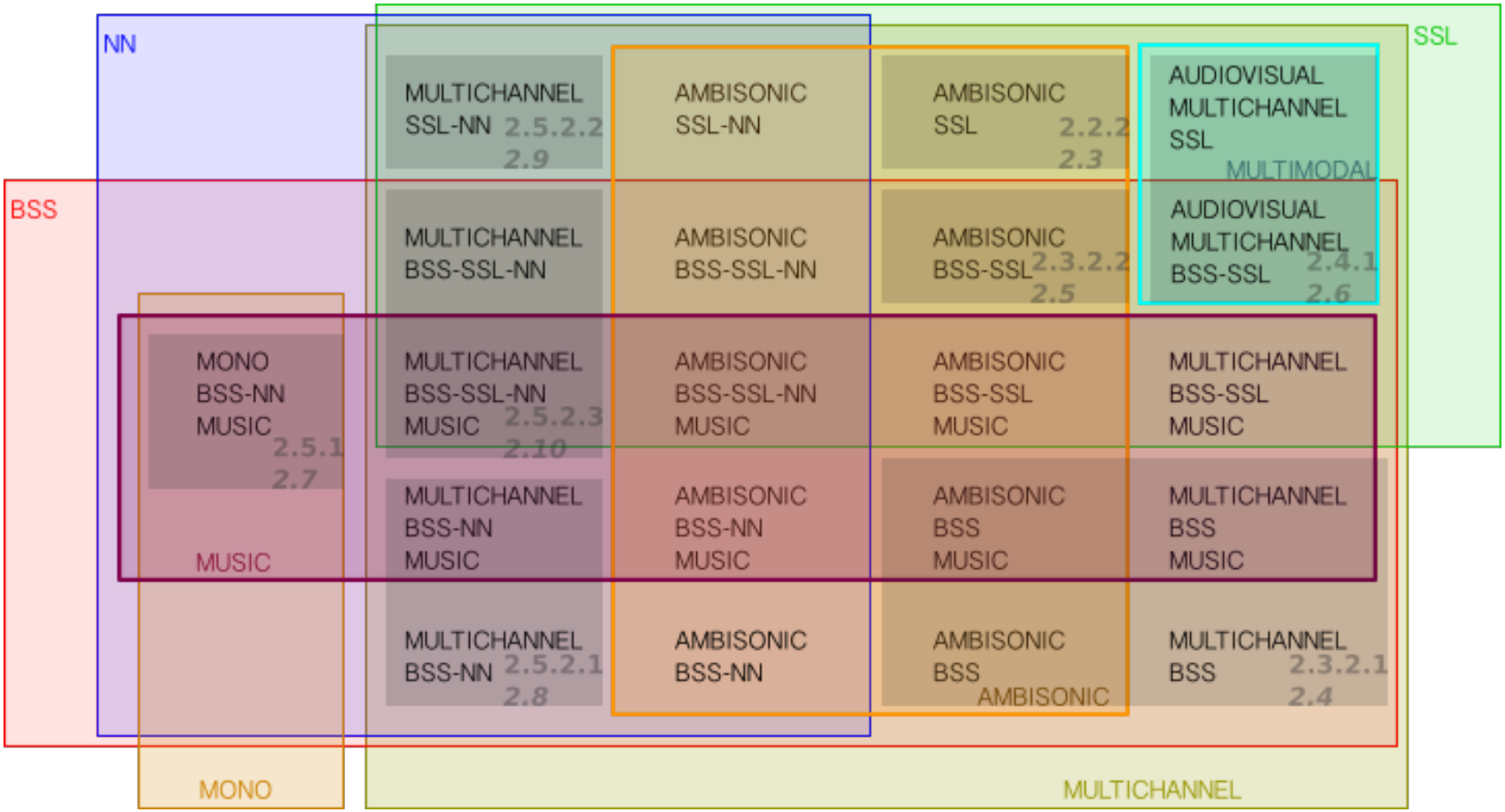
Summary

Euler Diagram



Summary

Euler Diagram



Outline for Section 3

Goals & Contributions

Conclusions

- BSS-SSL: stablished, but mostly horizontal FOA and speech

[1] A. Gannot *et. al.*, *A consolidated perspective on multi-microphone speech enhancement and source separation*, 2017

Goals & Contributions

Conclusions

- BSS-SSL: stablished, but mostly horizontal FOA and speech
- DNN: promising results for SSL/BSS, but mostly mic arrays and speech

[1] A. Gannot *et. al.*, *A consolidated perspective on multi-microphone speech enhancement and source separation*, 2017

Goals & Contributions

Conclusions

- BSS-SSL: stablished, but mostly horizontal FOA and speech
- DNN: promising results for SSL/BSS, but mostly mic arrays and speech
- Multimodal: *"the area of audio-visual speech processing remains largely understudied despite its great promise"*[1]

[1] A. Gannot *et. al.*, *A consolidated perspective on multi-microphone speech enhancement and source separation*, 2017

Goals & Contributions

Conclusions

BSS Guidelines[2]:

1. Consider number of sources and microphones

[2] A. Gannot *et. al.*, *A consolidated perspective on multi-microphone speech enhancement and source separation*, 2017

Goals & Contributions

Conclusions

BSS Guidelines[2]:

1. Consider number of sources and microphones
2. Exploit microphone array geometry

[2] A. Gannot *et. al.*, *A consolidated perspective on multi-microphone speech enhancement and source separation*, 2017

Goals & Contributions

Conclusions

BSS Guidelines[2]:

1. Consider number of sources and microphones
2. Exploit microphone array geometry
3. Exploit prior/additional information

[2] A. Gannot *et. al.*, *A consolidated perspective on multi-microphone speech enhancement and source separation*, 2017

Goals & Contributions

Goal

Research Goal:

- Investigation, adaptation and improvement of existing algorithms of Blind Source Separation for application to Ambisonics, specially focusing on musical applications.

Goals & Contributions

Contributions

Collateral Contributions:

1. Investigate SSL for HOA based on DNNs

Goals & Contributions

Contributions

Collateral Contributions:

1. Investigate SSL for HOA based on DNNs
2. Apply *Contribution I* to BSS, focusing on music

Goals & Contributions

Contributions

Collateral Contributions:

1. Investigate SSL for HOA based on DNNs
2. Apply *Contribution I* to BSS, focusing on music
3. Investigate raw multichannel BSS for HOA based on DNNs

Goals & Contributions

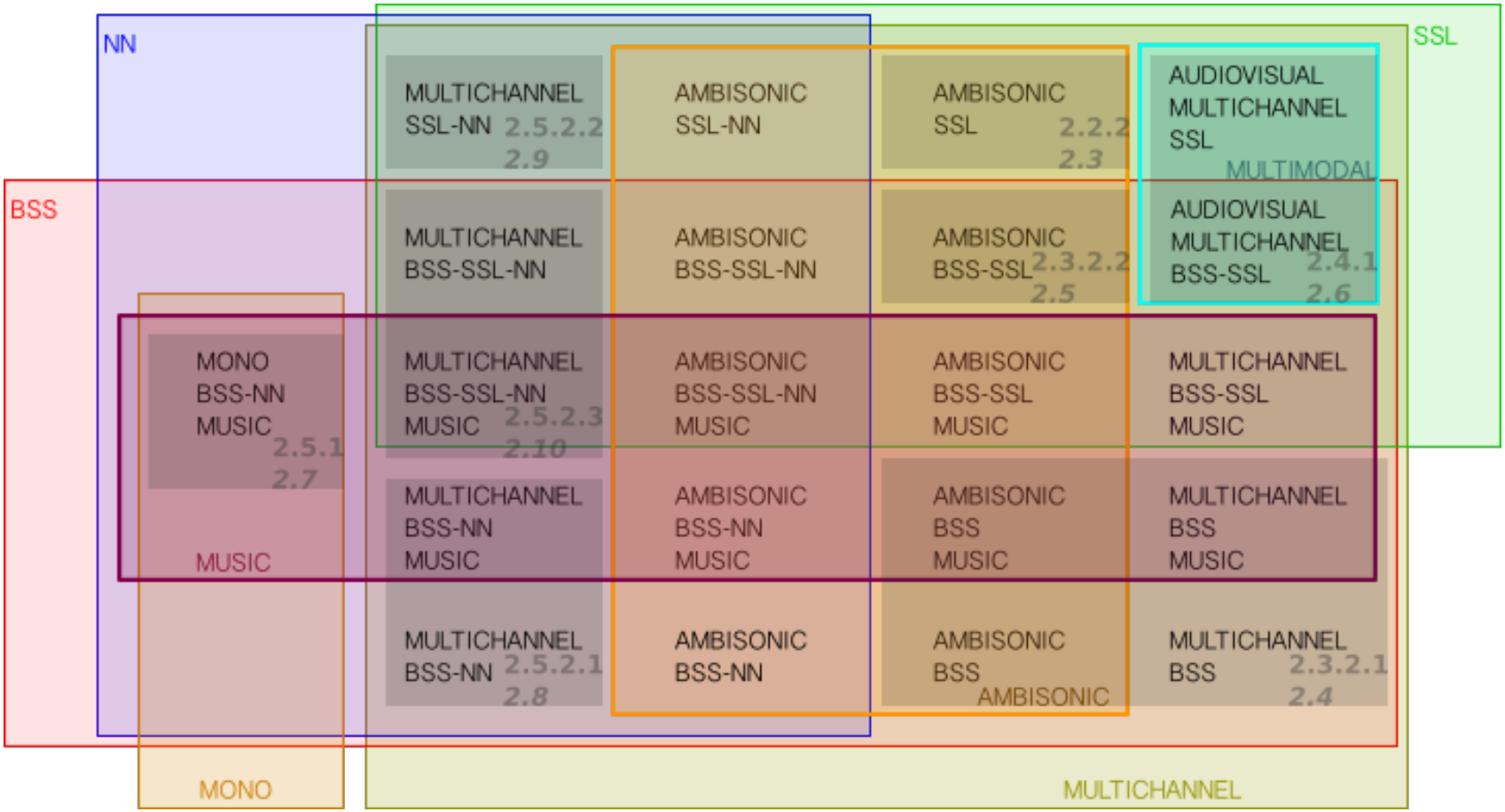
Contributions

Collateral Contributions:

1. Investigate SSL for HOA based on DNNs
2. Apply *Contribution I* to BSS, focusing on music
3. Investigate raw multichannel BSS for HOA based on DNNs
4. New approach to multimodal BSS from immersive audiovisual content

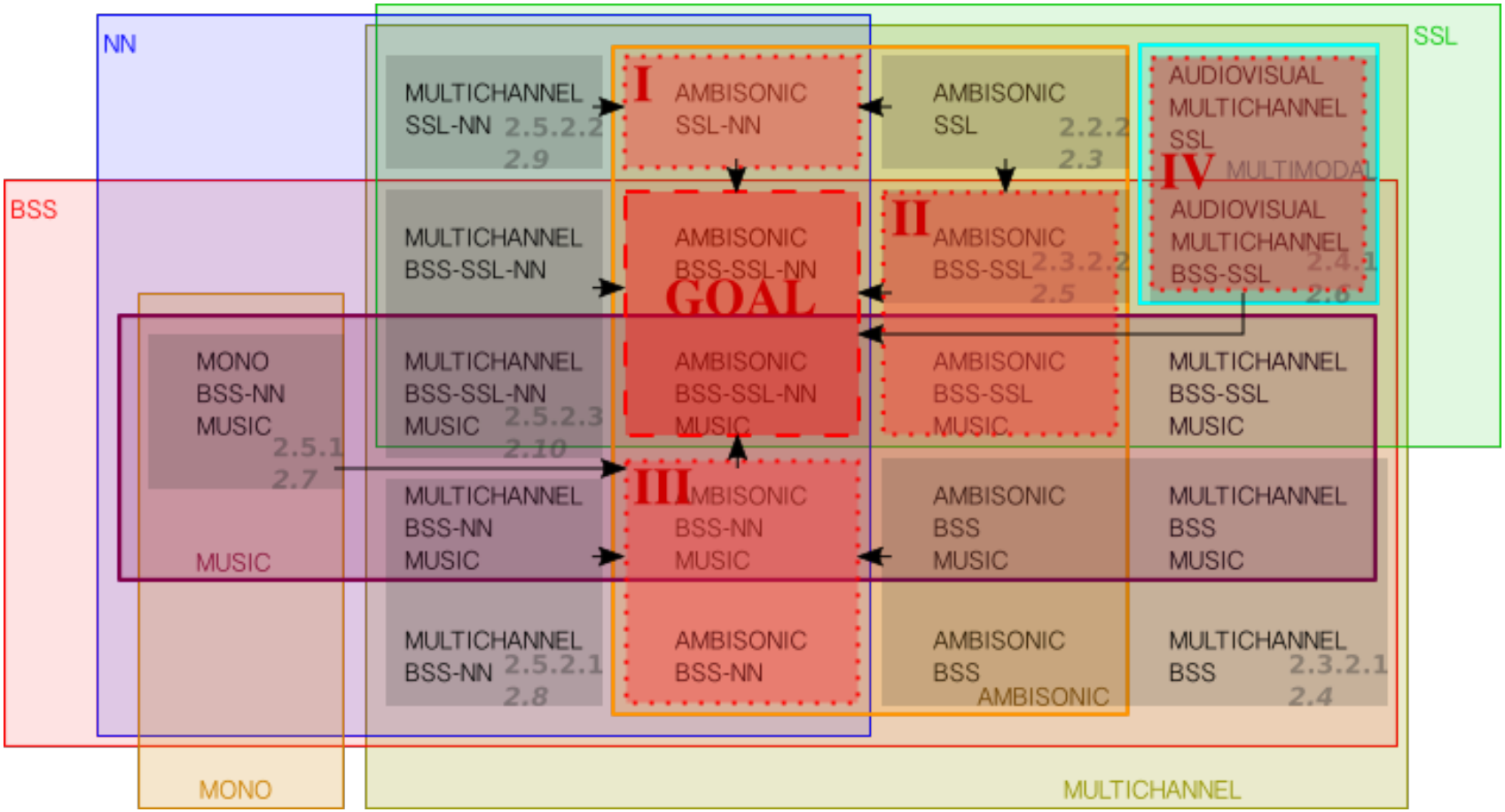
Goals & Contributions

Contributions



Goals & Contributions

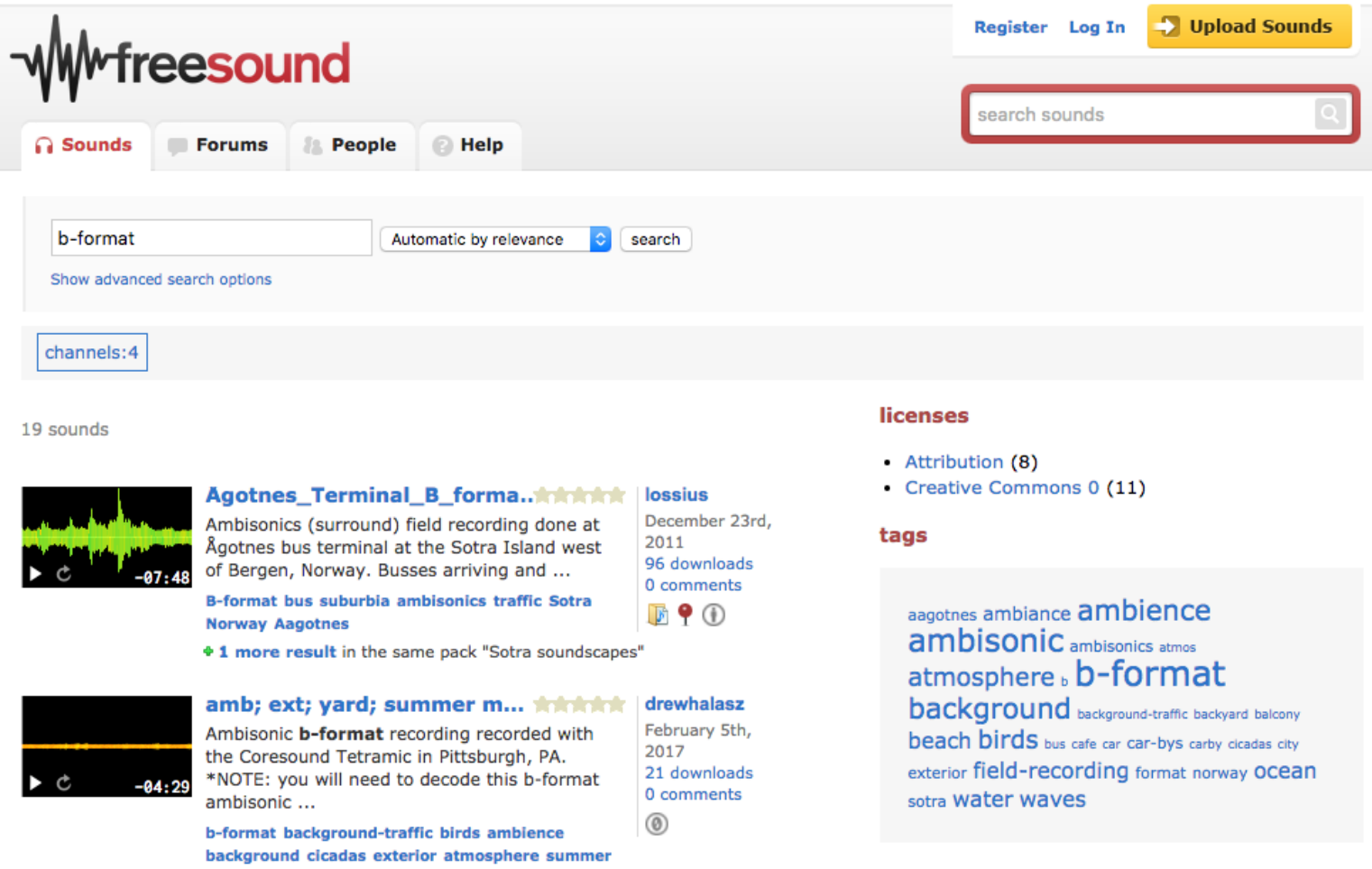
Contributions



Outline for Section 3

Methodology

Dataset



Methodology

Dataset

Music:

- MSD100/DSD100
- MIR-1k

Speech:

- TIMIT
- WSJCAM0

Methodology

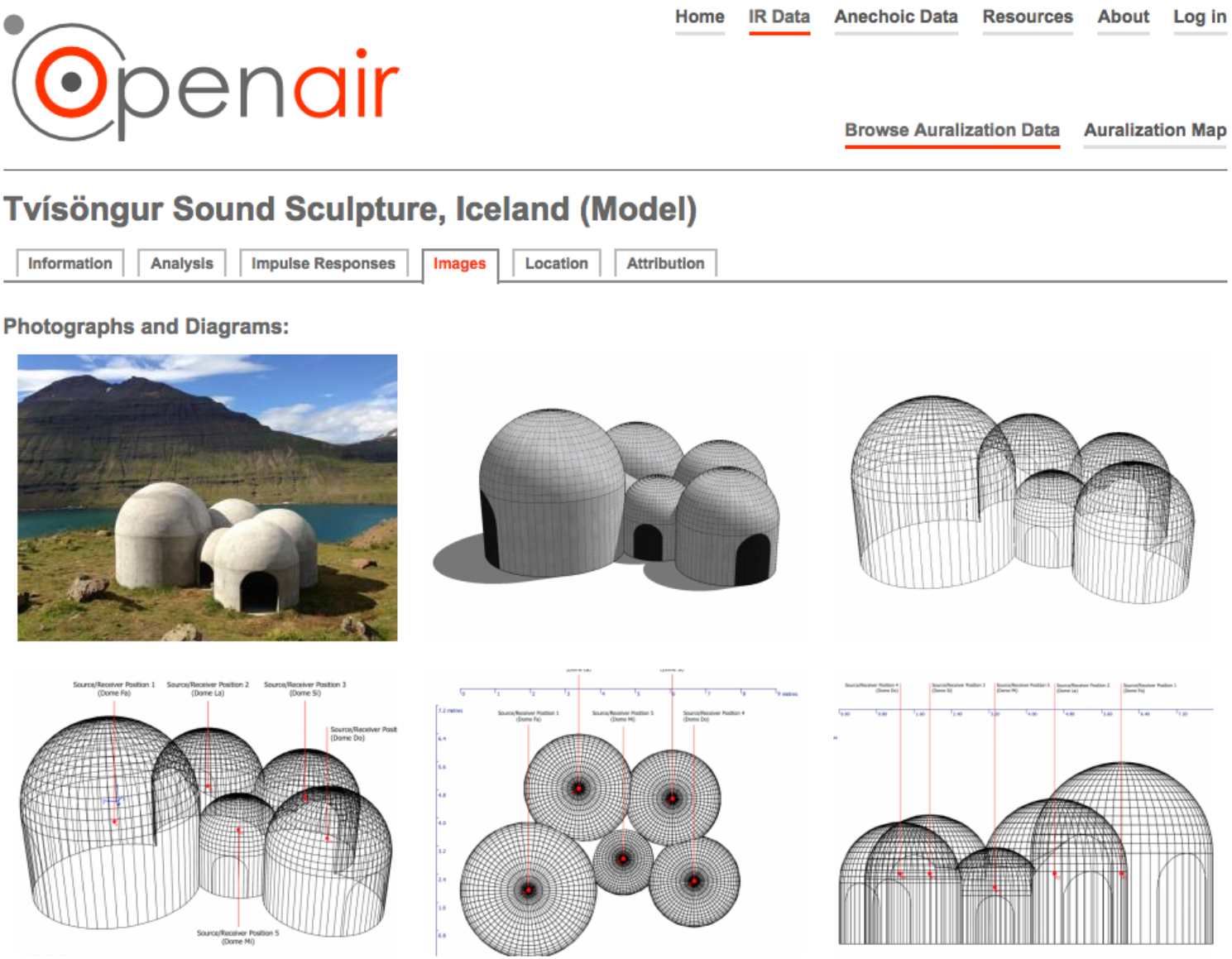
Dataset

Ambisonics Impulse Response:

- SMIR Generator
- OpenAirLib

Methodology

Dataset



Methodology

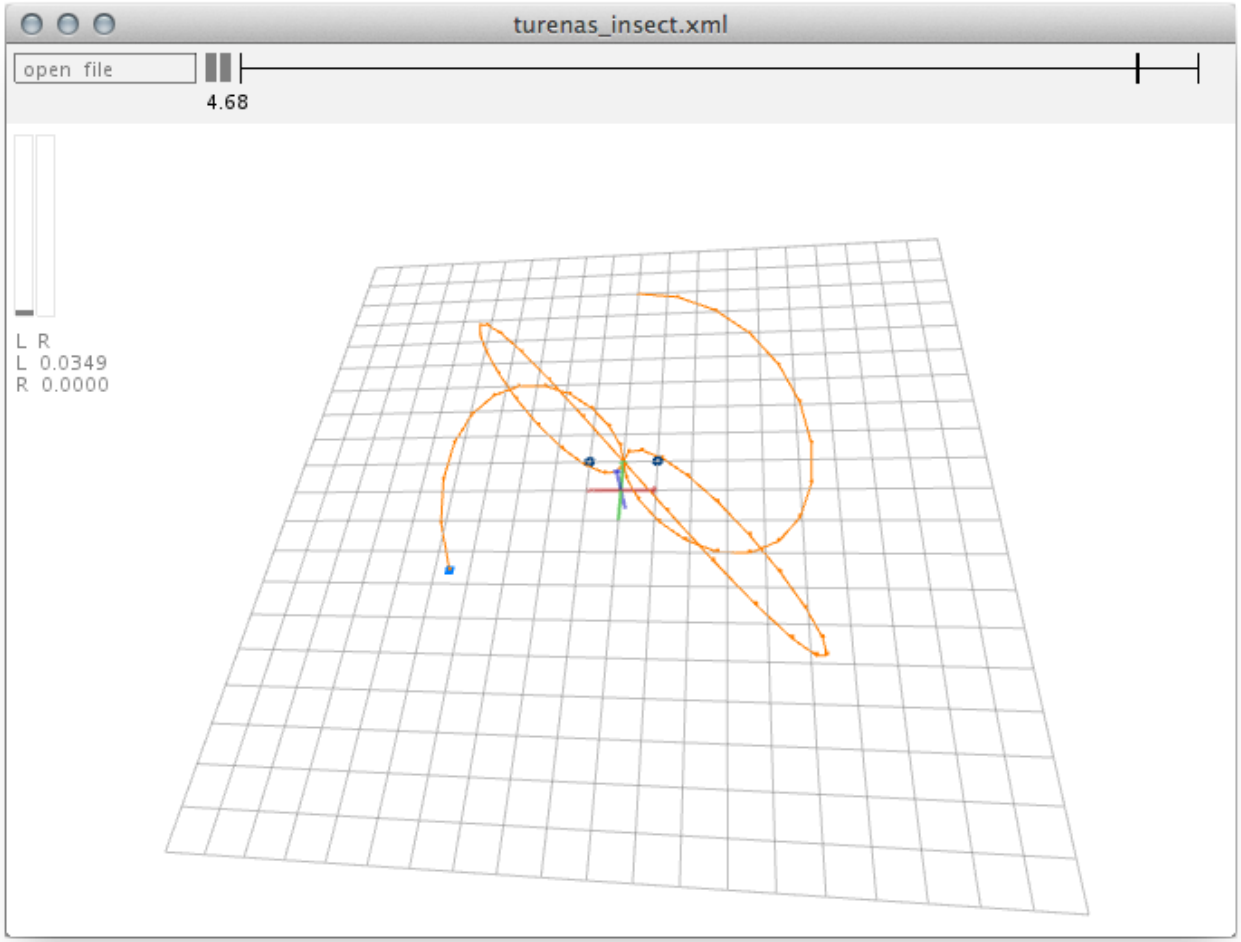
Dataset Contributions

Collateral Contributions:

1. Investigate SSL for HOA based on DNNs
2. Apply *Contribution I* to BSS, focusing on music
3. Investigate raw multichannel BSS for HOA based on DNNs
4. New approach to multimodal BSS from immersive audiovisual content
5. New tool for procedural creation of reverberant sound scenes, for training and evaluation purposes

Methodology

Scene Description: SpatDIF



Methodology

Experimental Setups

Ambisonics Microphones Availability:

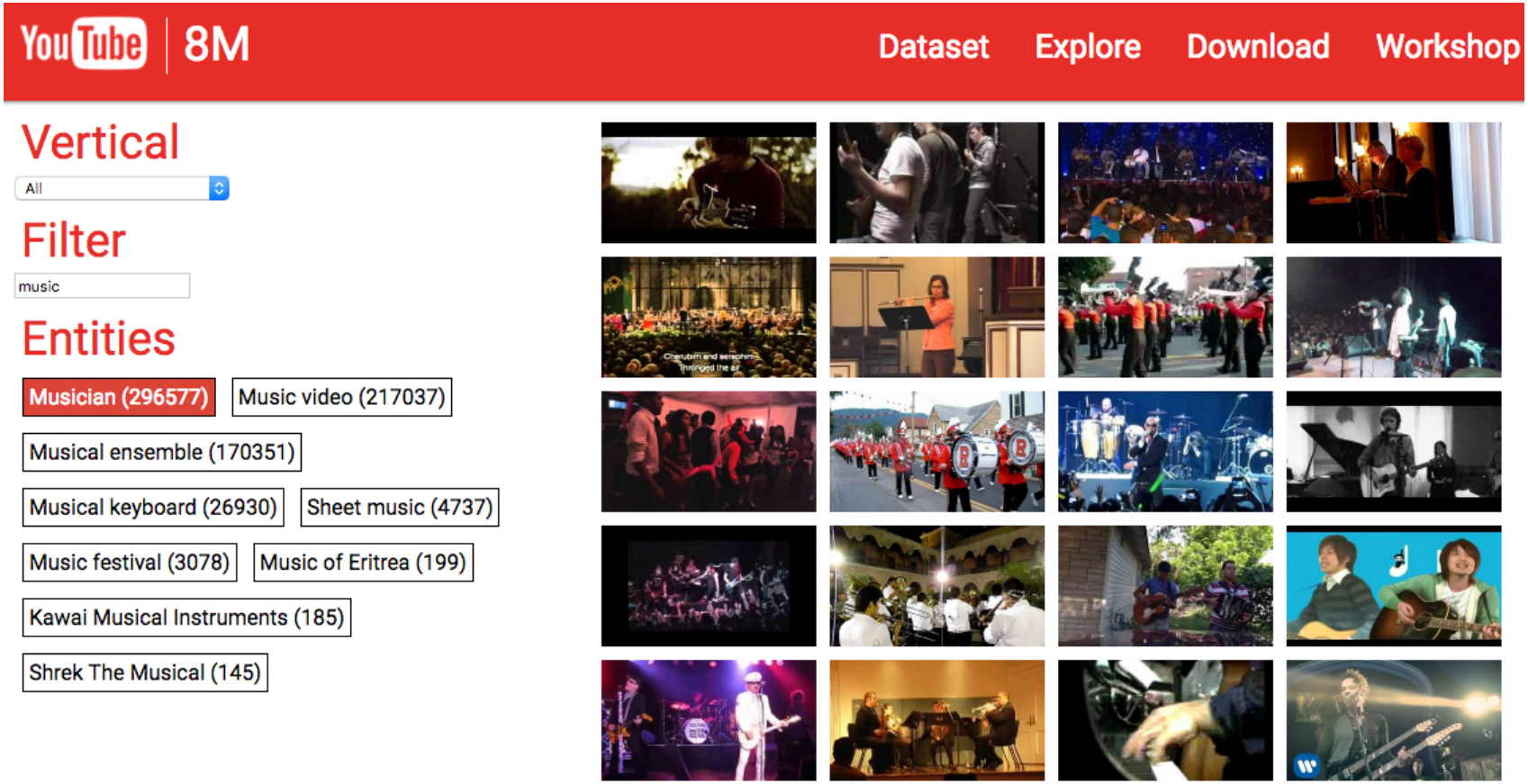
- SoundField SPS422B
- EigenMike
- Ambeo
- Zoom H2n

Methodology

Experimental Setups

Methodology

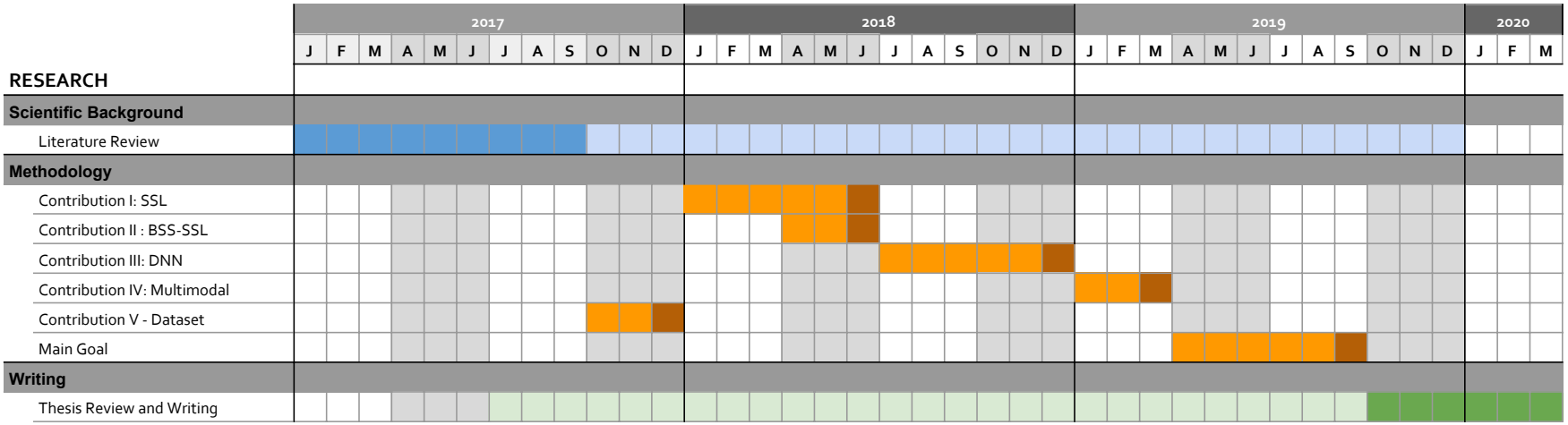
Experimental Setups



Outline for Section 3

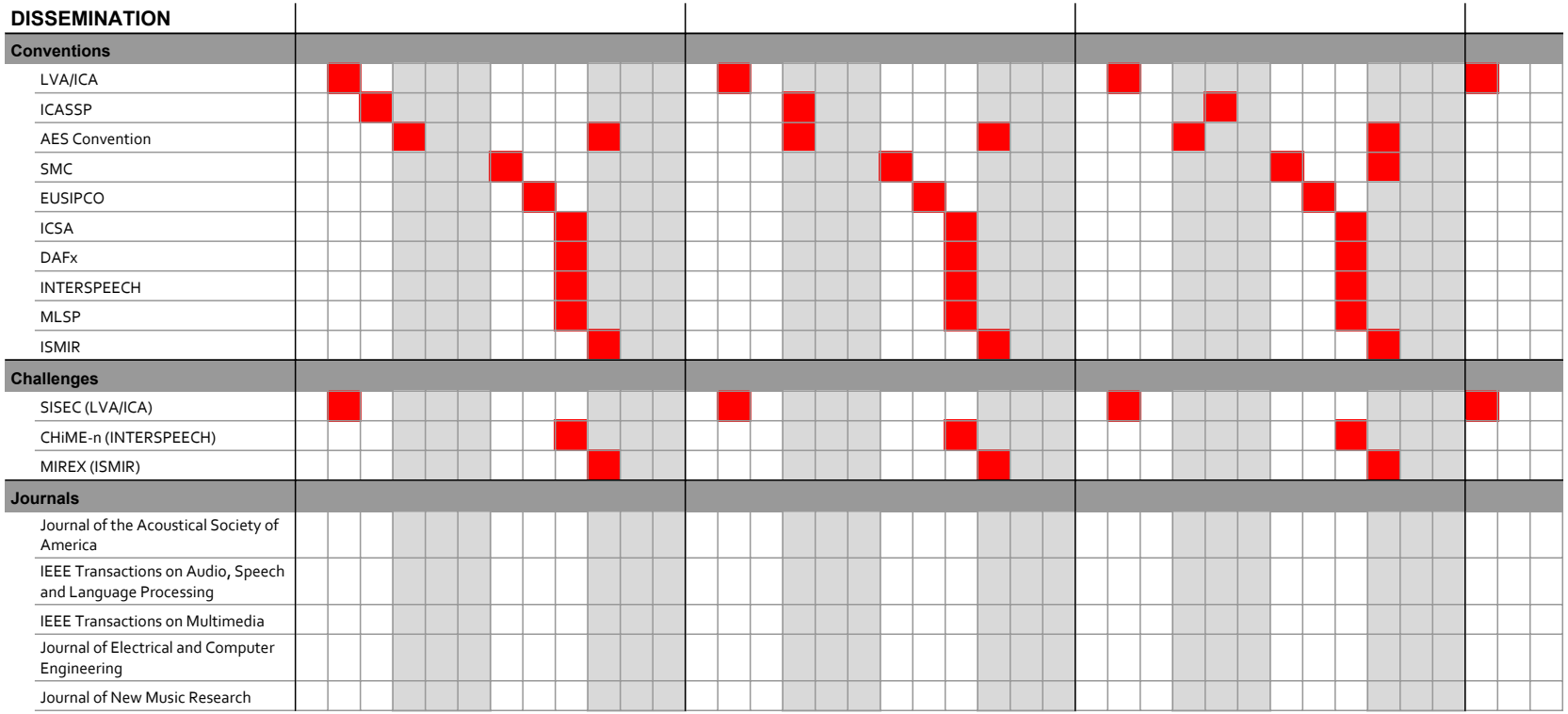
Schedule & Dissemination

Schedule



Schedule & Dissemination

Dissemination



Thank you

Questions?