VIBE PROJECT

Virtual Biomedical and STEM/STEAM Education

2021-1-HU01-KA220-HED-000032251

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

PÉCSI TUDOMÁNYEGYETEM UNIVERSITY OF PÉCS | U.PORTO | Politechnika Śląska | DEX innovation centre

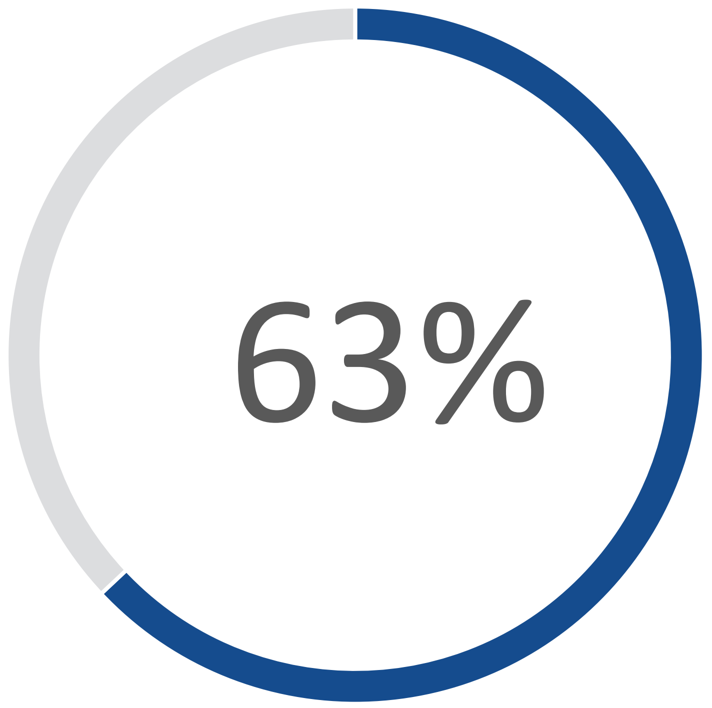
VIBE PROJECT

PATTERN RECOGNITION IN BIOMEDICAL ENGINEERING

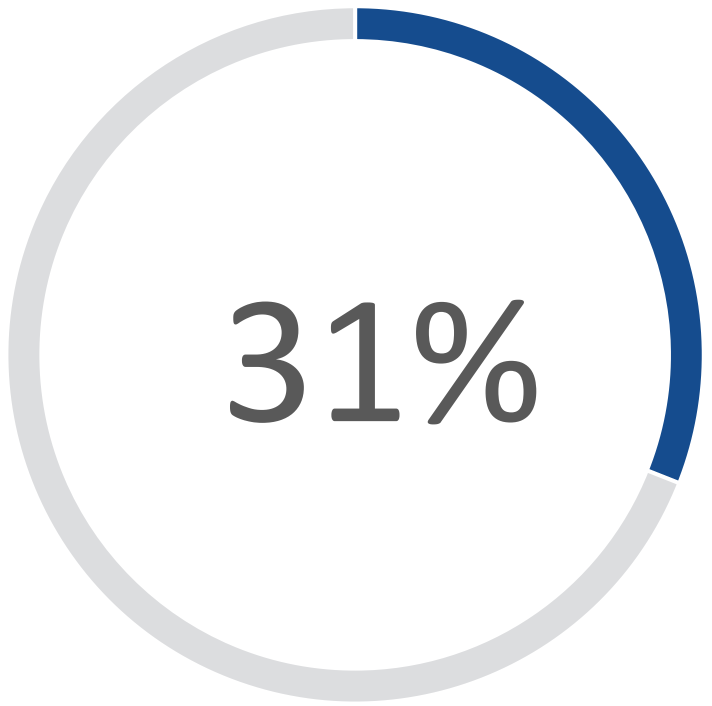
ANALYSIS OF SPATIAL ACOUSTIC MODELS OF SIBILANT SOUNDS IN SIGMATISM DIAGNOSIS

Speech disorders in Poland

3



Dyslalia among children
in Poland,



including sigmatism
(lisp).

Elżbieta M. Minczakiewicz, Dyslalia in comparison with other speech defects and disorders in preschool and school-age children, PAN, Warsaw, 2017.

Sigmatism

A lisp (sigmatism) is a speech impairment in which a person misarticulates sibilants (in Polish):

/s,z,ts,dz/

/ʃ,ʒ,ʧ,ʤ/

/ɕ,ʑ,ʨ,ʥ/

www.Thinkstockphotos.com

Causes

- abnormalities in the structure and functioning of the articulation organs, anatomical changes,
- low motor efficiency of the articulation organs,
- incorrect structure and functioning of the hearing organ.

Types of articulations*

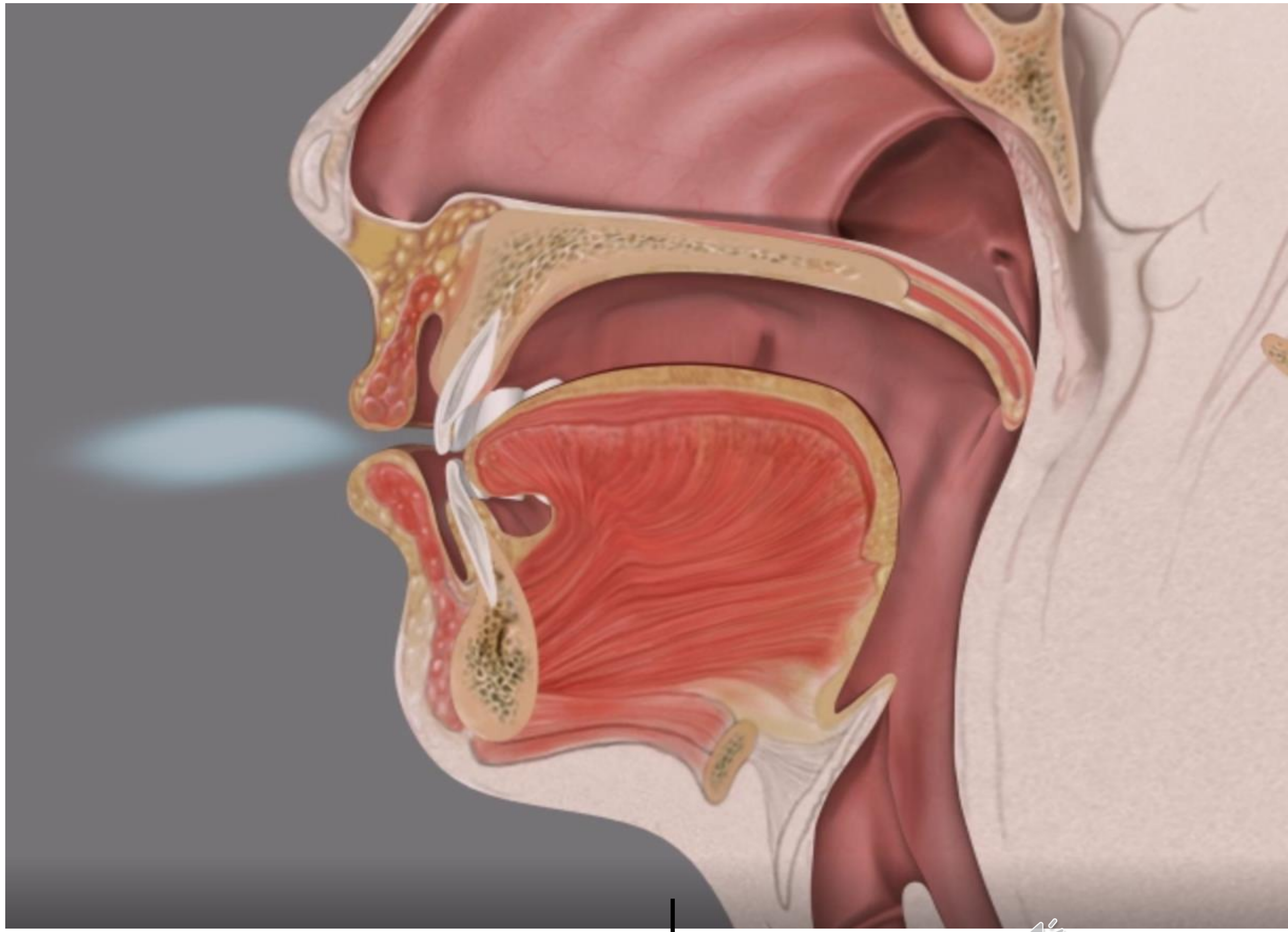
Normative articulation

Interdental articulation

Addental articulation

Dental articulation

5



Normative realization of the sound /s/

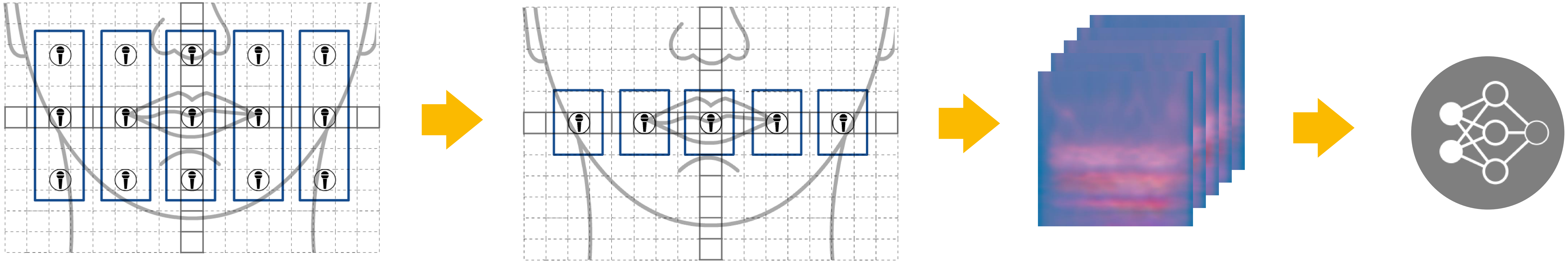
*due to the location of the apex in the oral cavity

The aim of the research conducted

MAIN AIM	DEVELOPMENT OF SPATIAL ACOUSTIC MODELS OF SIBILANTS AND THEIR USE IN THE CLASSIFICATION OF NORMATIVE AND NON-NORMATIVE SIBILANT ARTICULATIONS
SECONDARY AIM	DESIGN AND CONSTRUCTION OF A MEASUREMENT STATION ENABLING SPATIAL ACQUISITION OF THE SPEECH SIGNAL

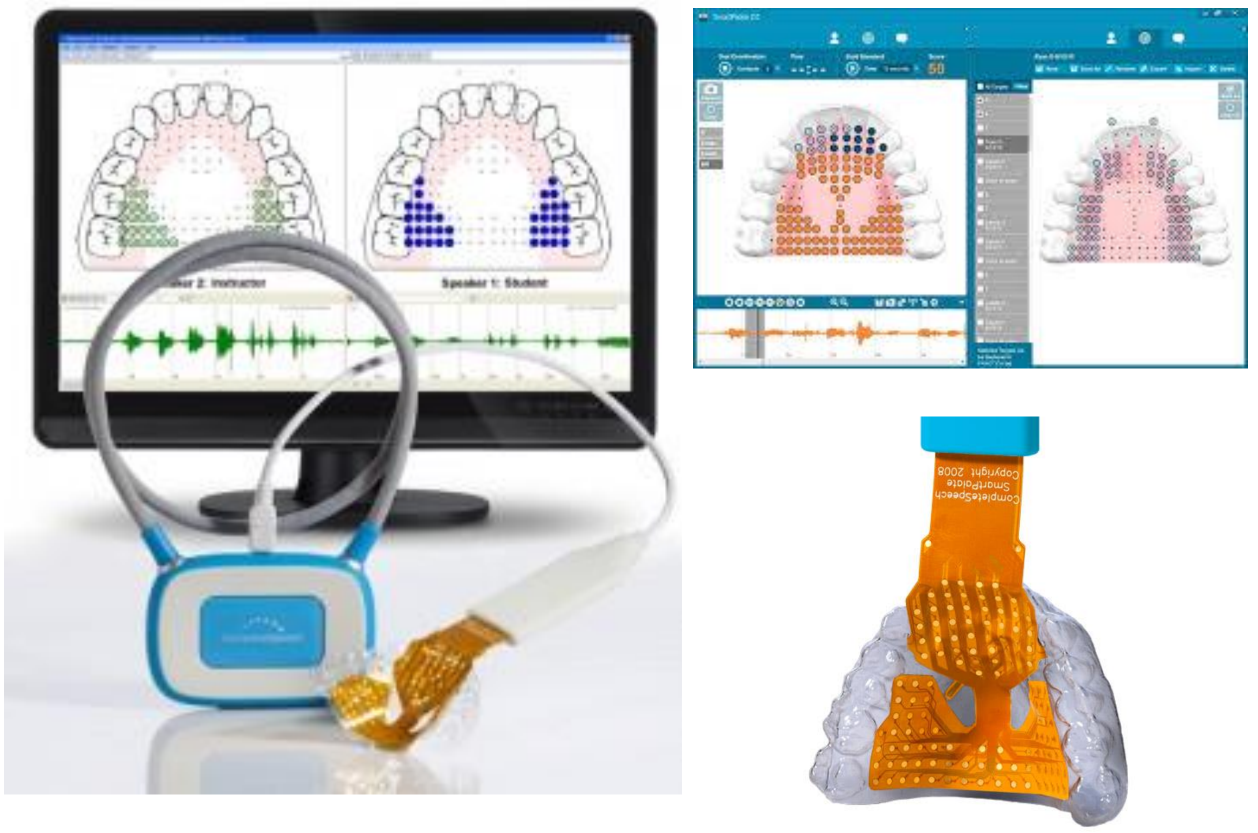
Research thesis

THE USE OF SPATIAL SPEECH SIGNAL PROCESSING METHODOLOGY ENABLES THE RECOGNITION OF SIGMATISM AND DETERMINATION OF ITS TYPE



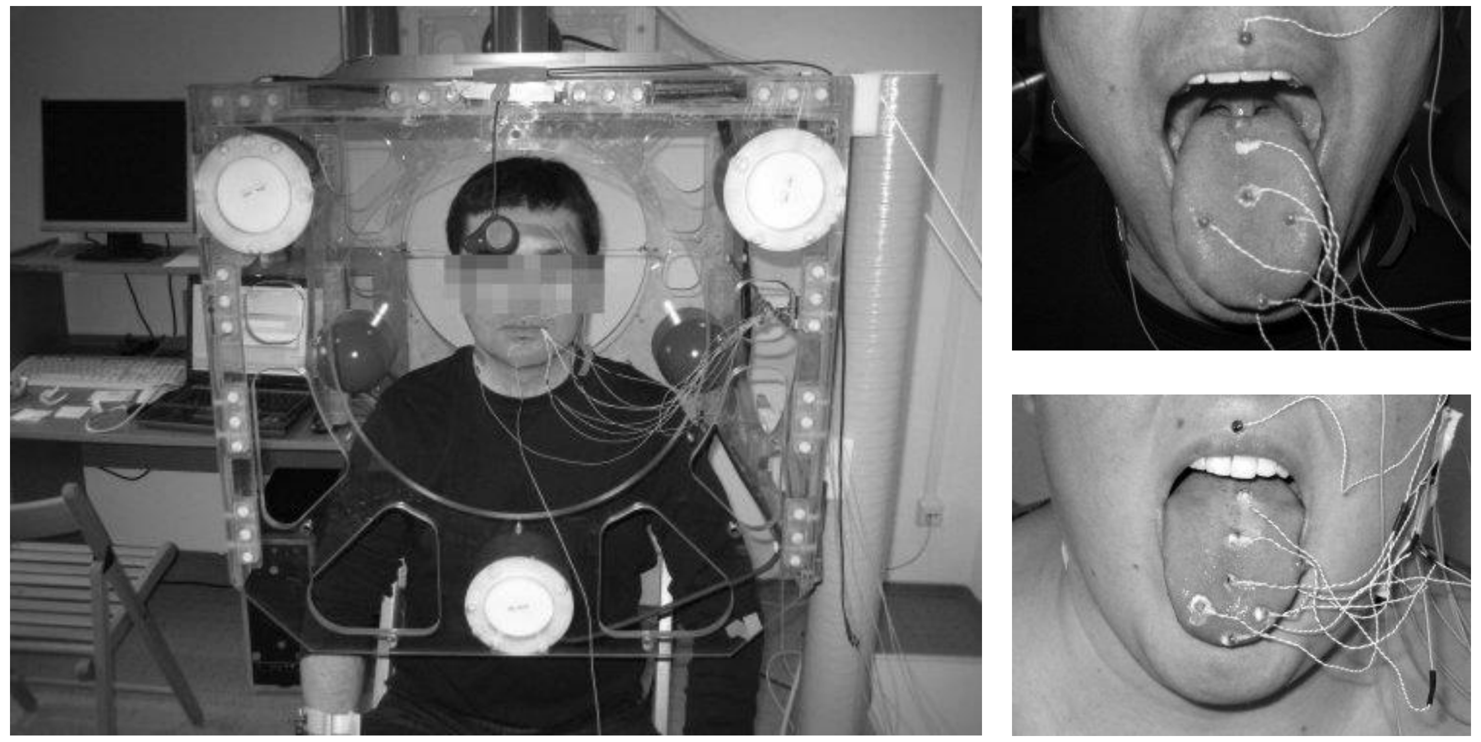
Measurement methods

Electropalatography



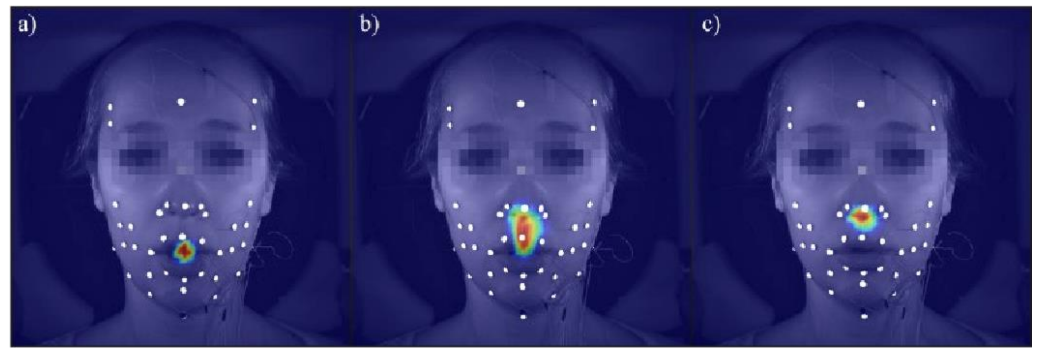
www.
completespeech.com/smartpalate/.pl

Electromagnetic
articulography



www.komlogo.pl

Acoustic
holography



Acoustic Field Distribution in Speech with the use of the Microphone Array

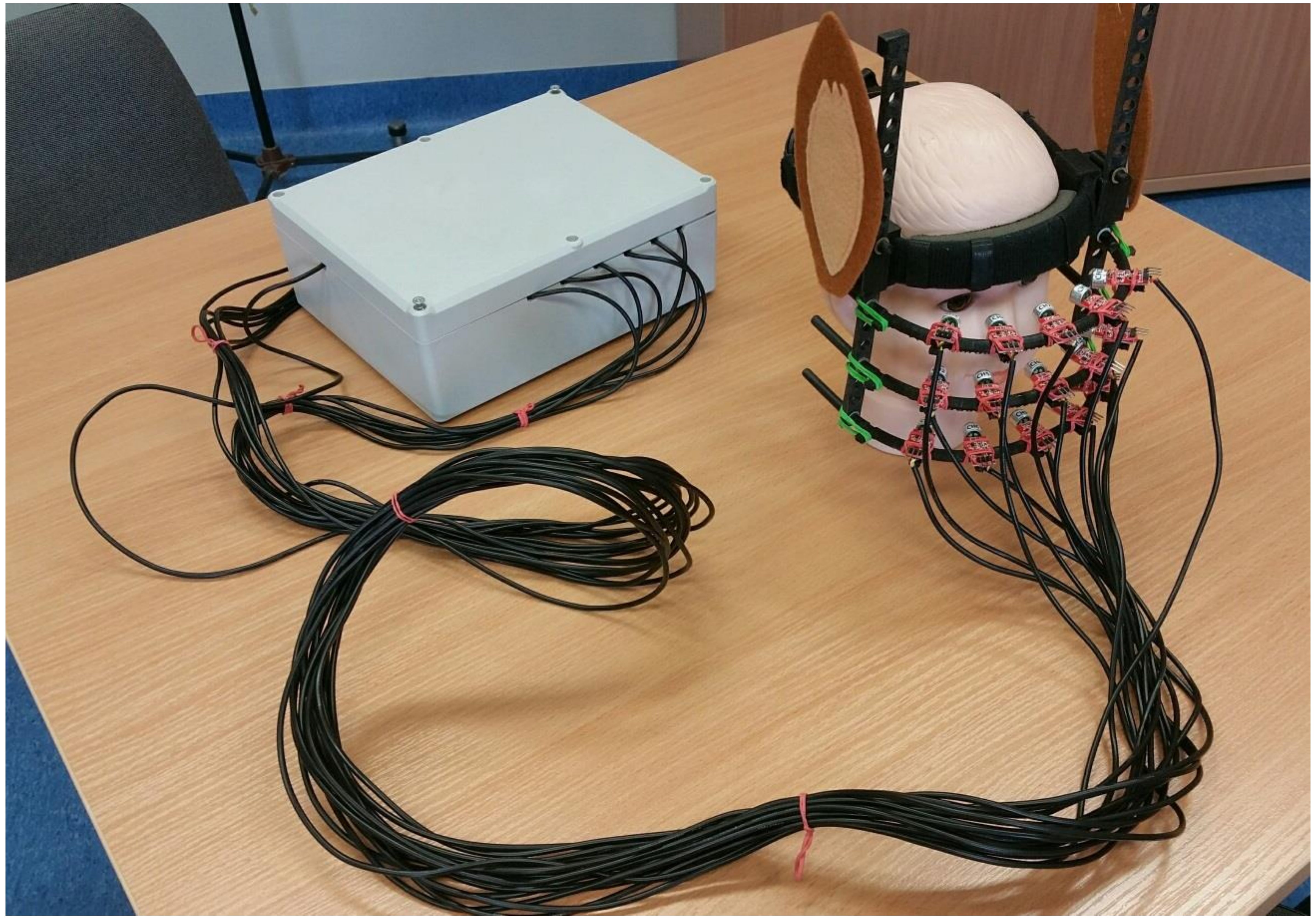
www.bksv.co
m

REQUIRED CRITERIA

Measuring station

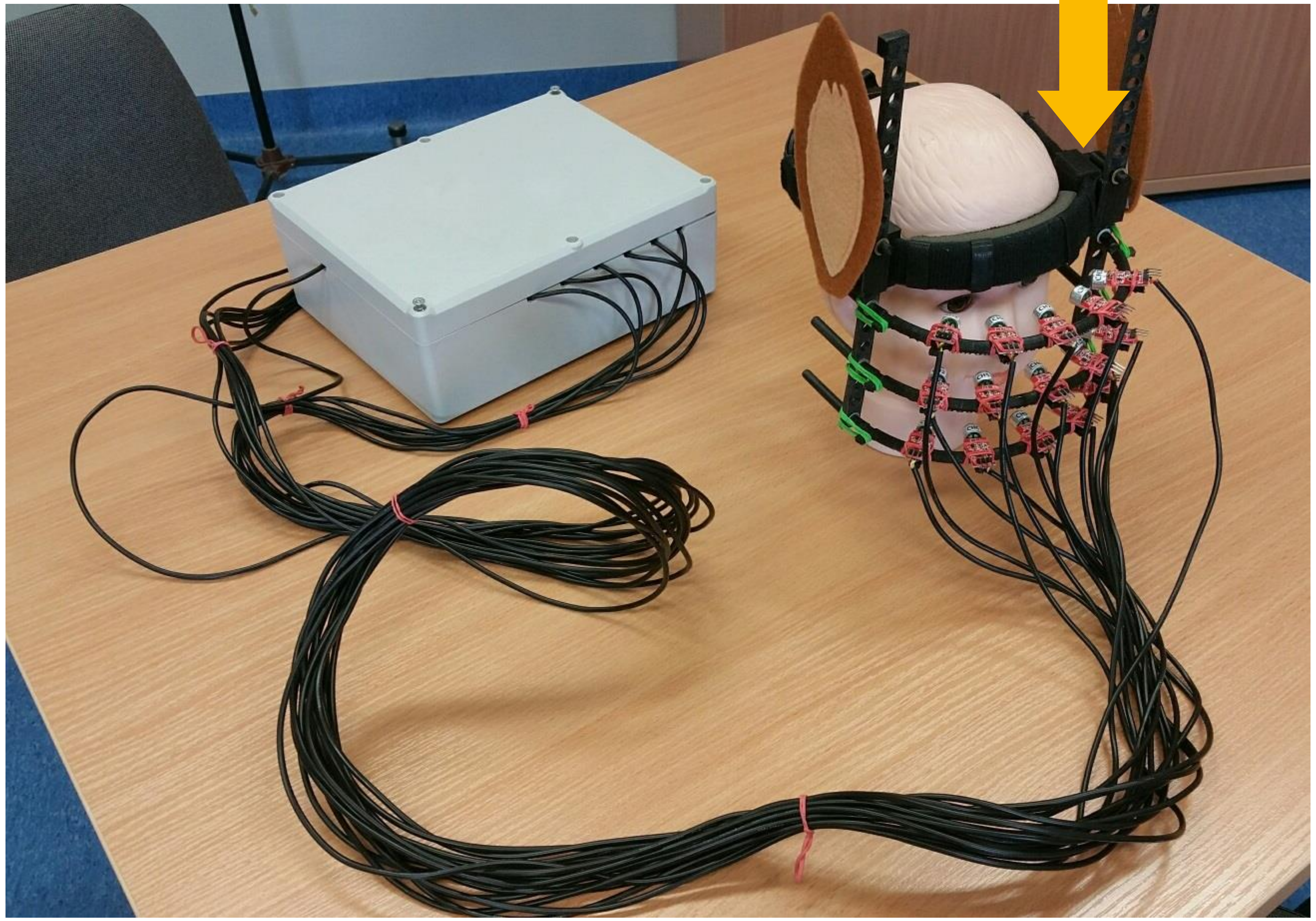
- multi-channel, spatial and repeatable sound recording,
- does not require a specialist measuring room
- access to raw data
- visually attractive for children
- no impact on the articulation and phonation process
- painless (non-invasive) data recording

Measuring station



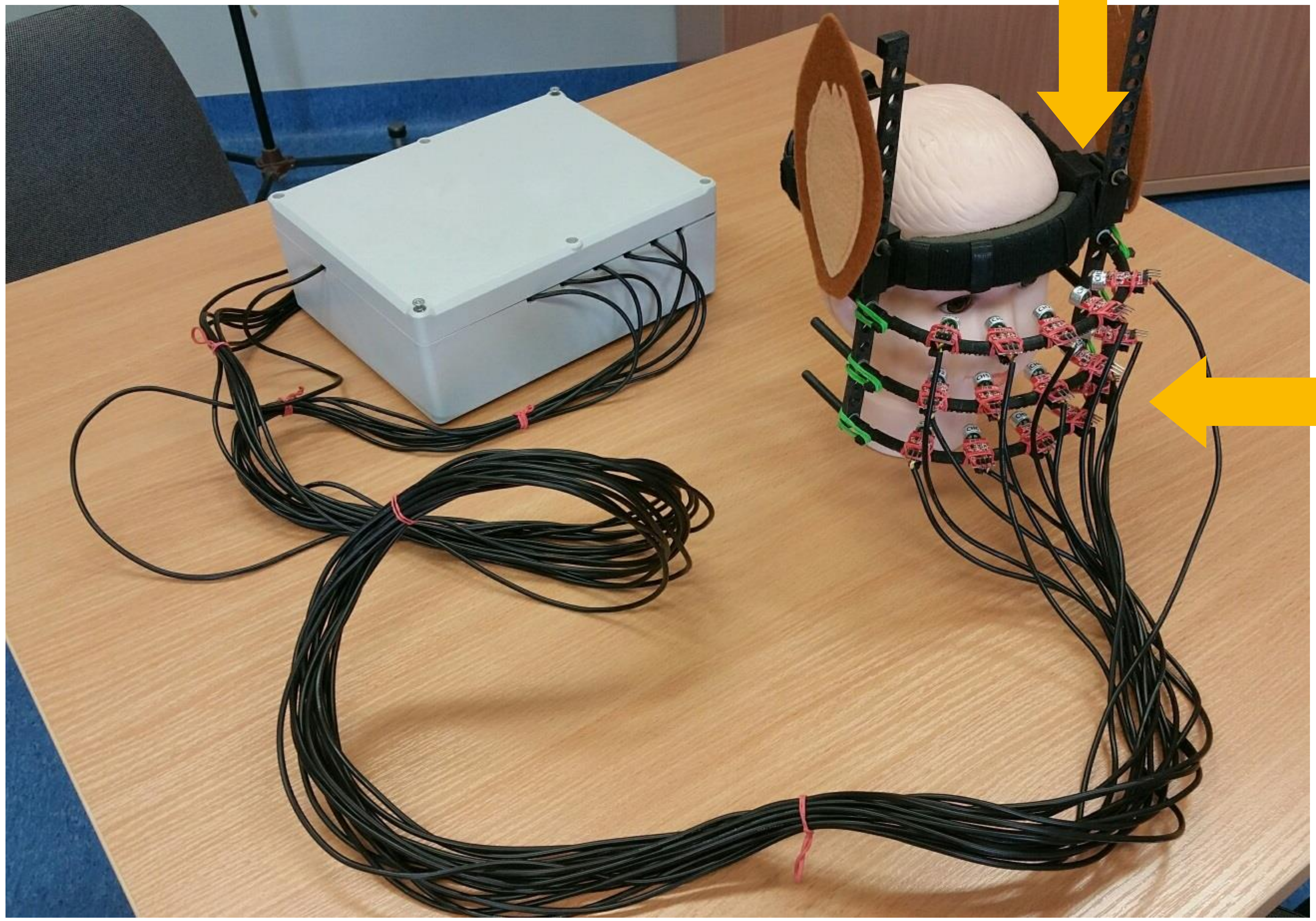
BUILDING A PROTOTYPE

Measuring station



Measuring station

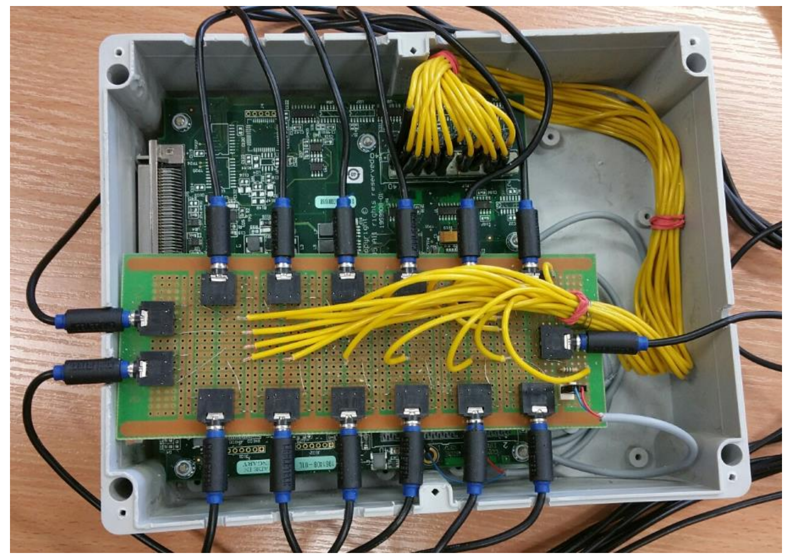




Sound pressure transducers WM-61a

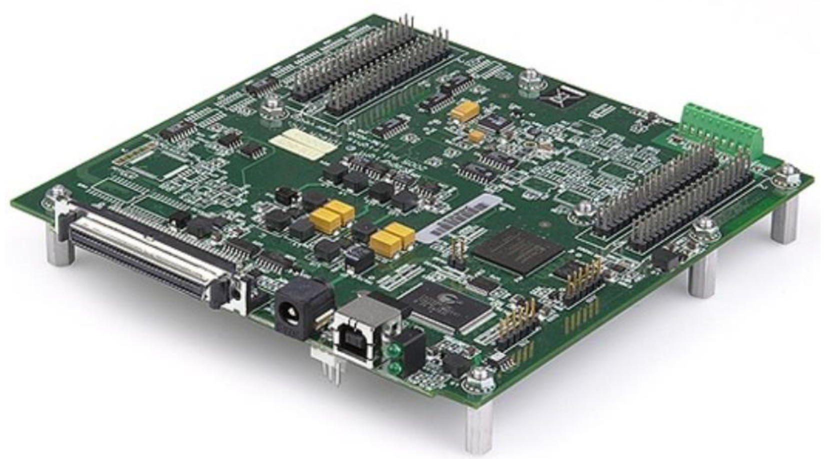
MAX9812 Microphone Preamplifier and Amplifier

Measuring station

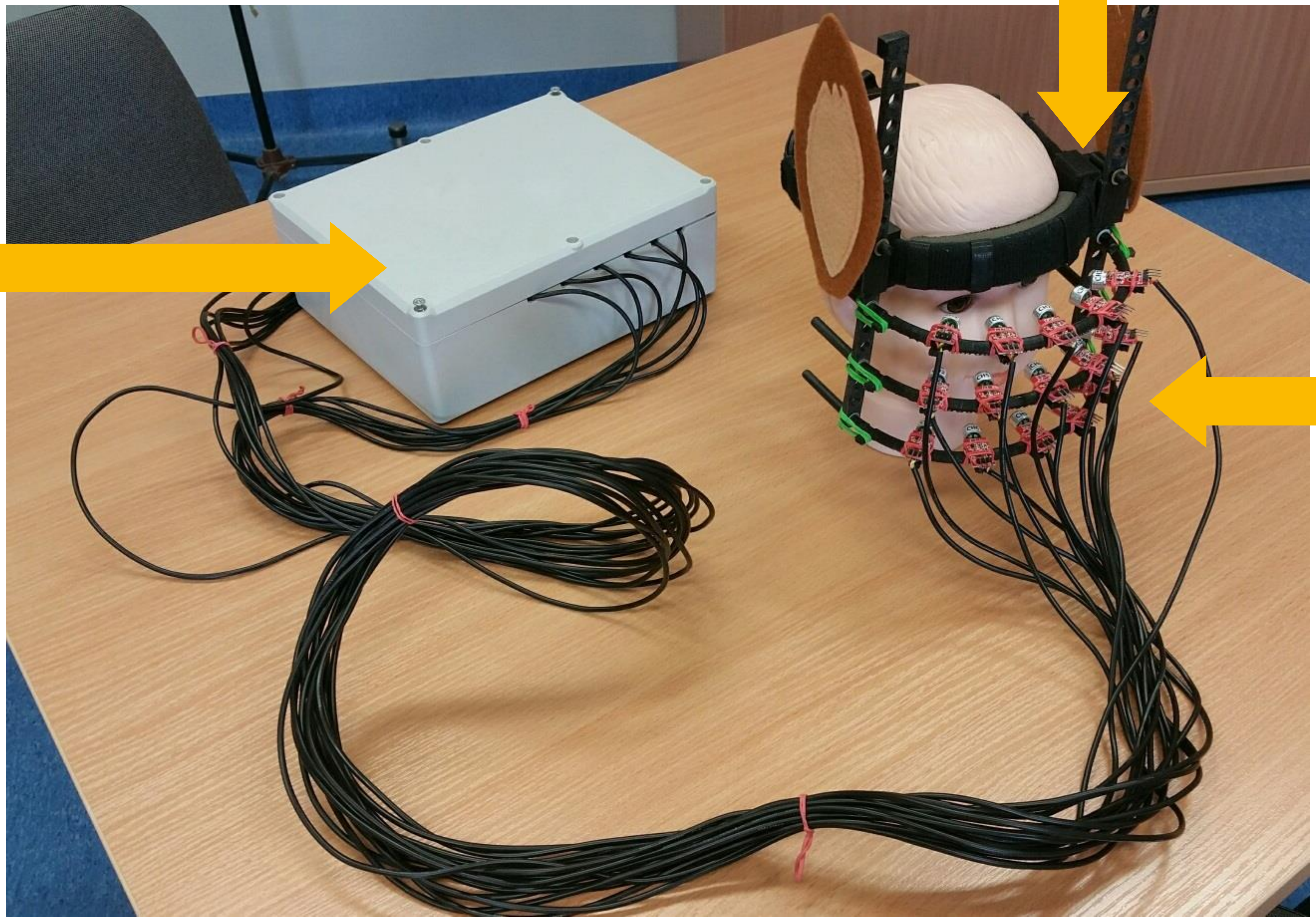


Microphone adapter

+



DaqBoard 3000USB Series Data Acquisition Card

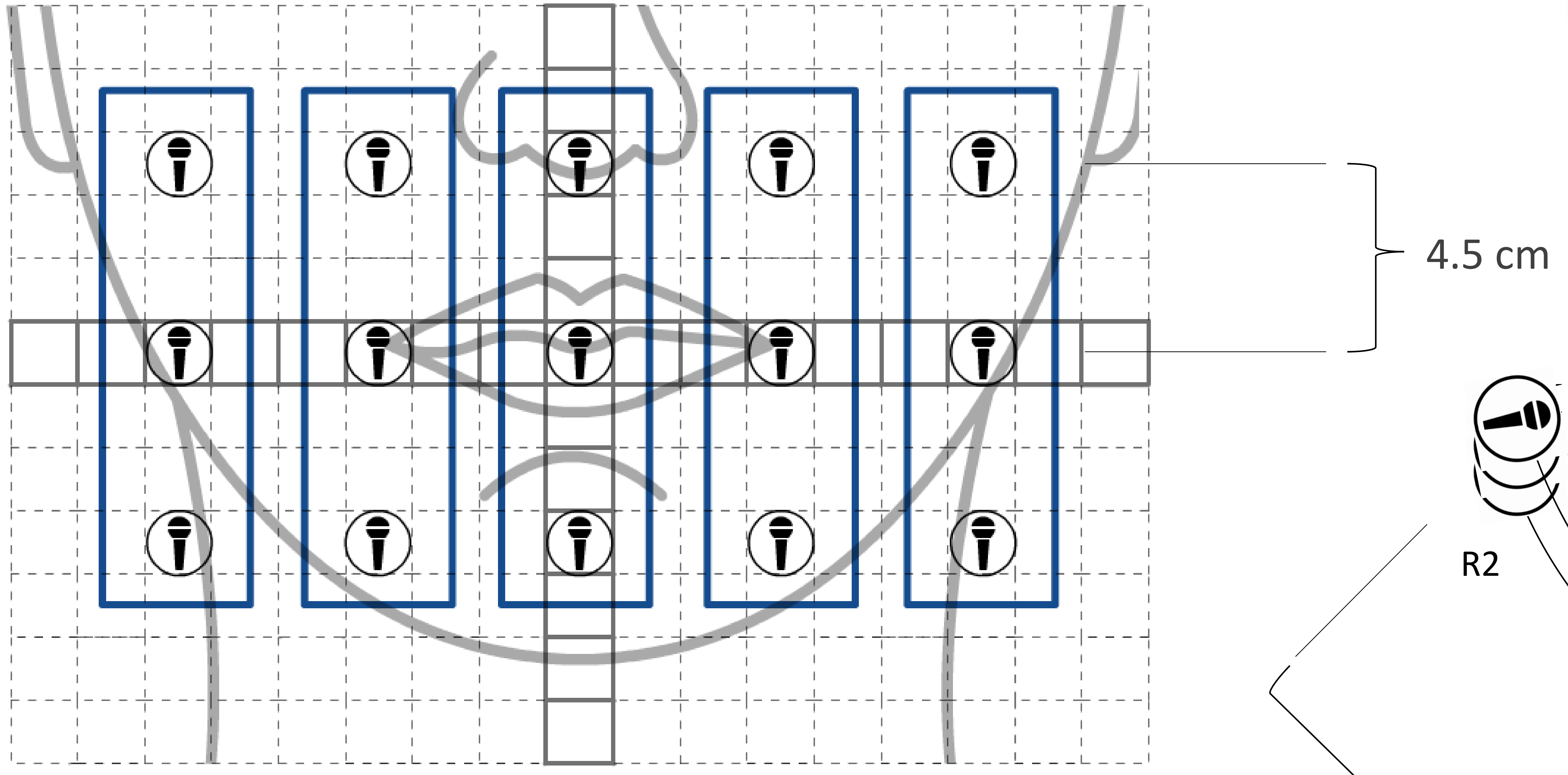


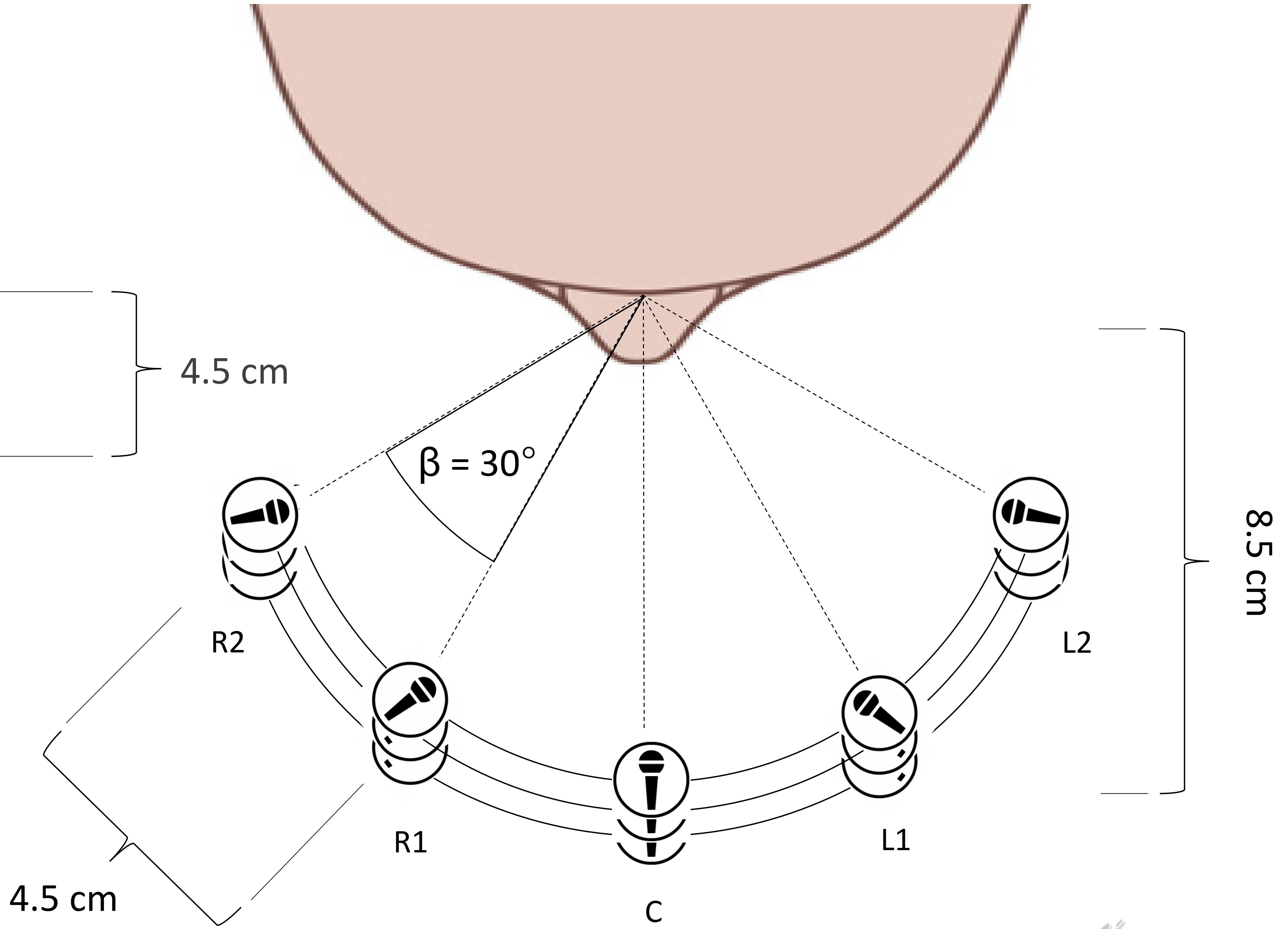
Sound pressure transducers WM-61a

+

MAX9812 Microphone Preamplifier and Amplifier

Measuring station





Measuring station

Number of channels	15
Number of groups	5
Distance from sound source	8,5 cm
Sampling rate	44,1 kHz

Microphone model	WM-61a
sensitivity	-35 dB
frequency range	20 Hz – 20 kHz
SNR	72 dB (65 dB)
THD	-72 dB (-65 dB)
power supply	2,3 V

Amplifier model	MAX9812
Gain	20 dB

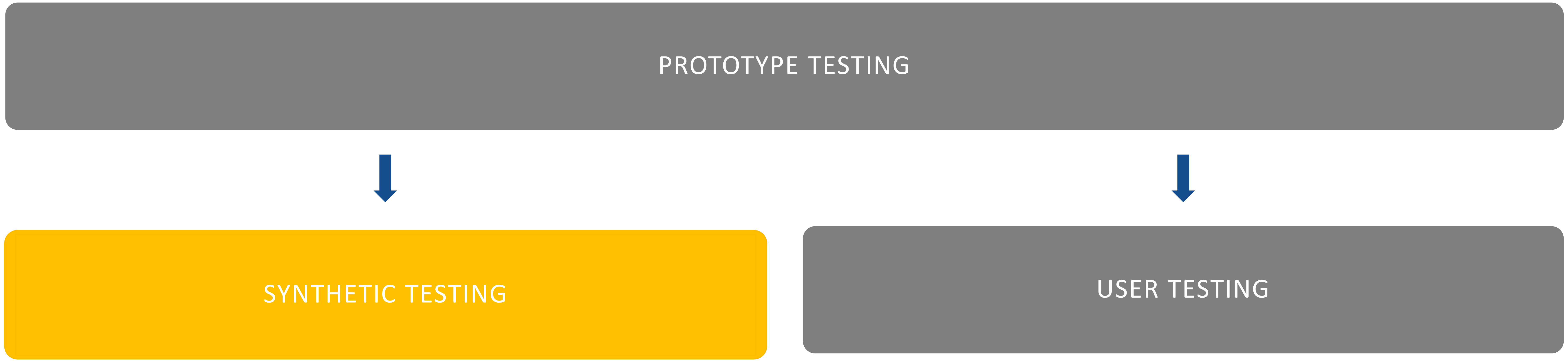
Measuring station

PROTOTYPE TESTING

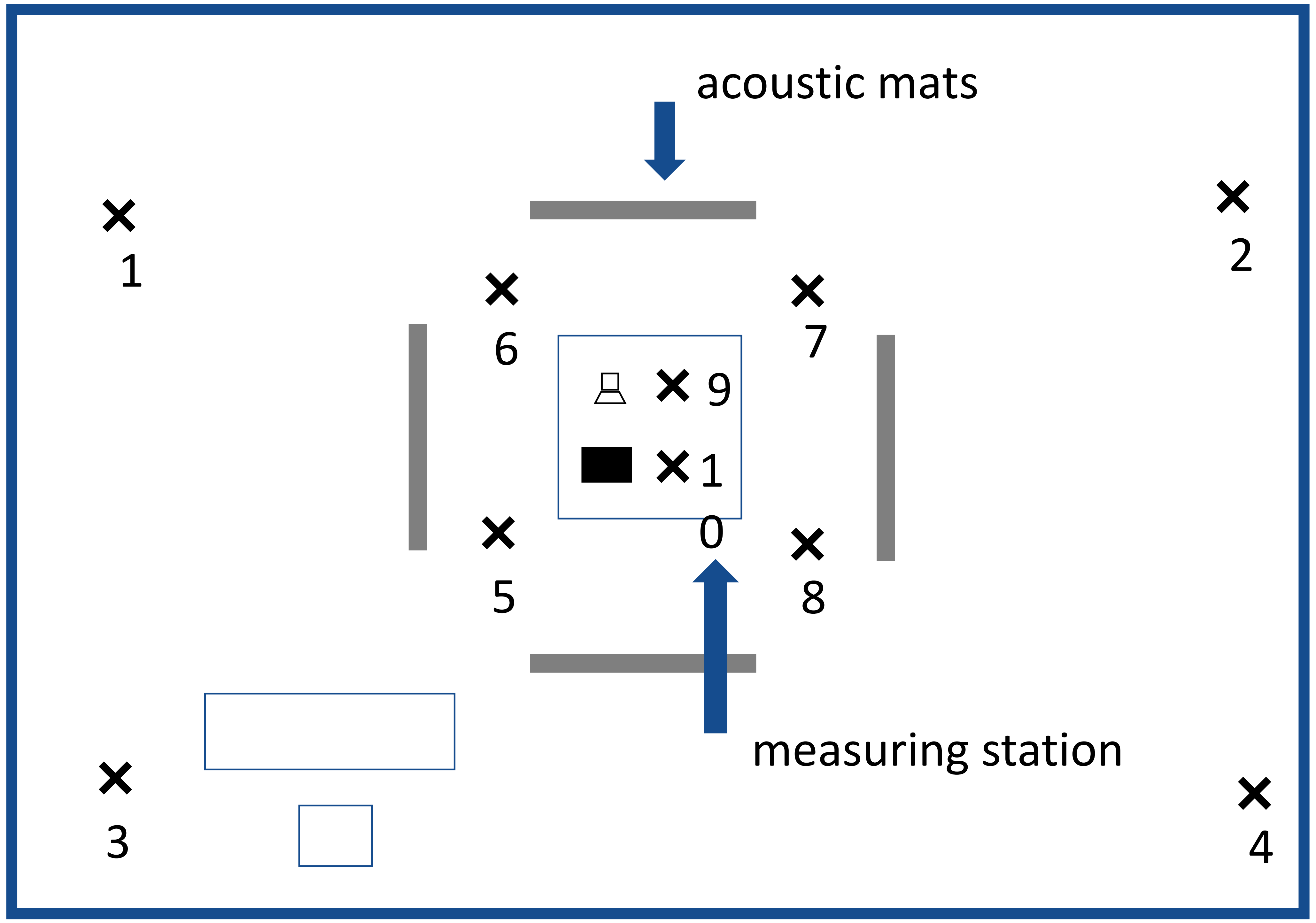
SYNTHETIC TESTING

USER TESTING

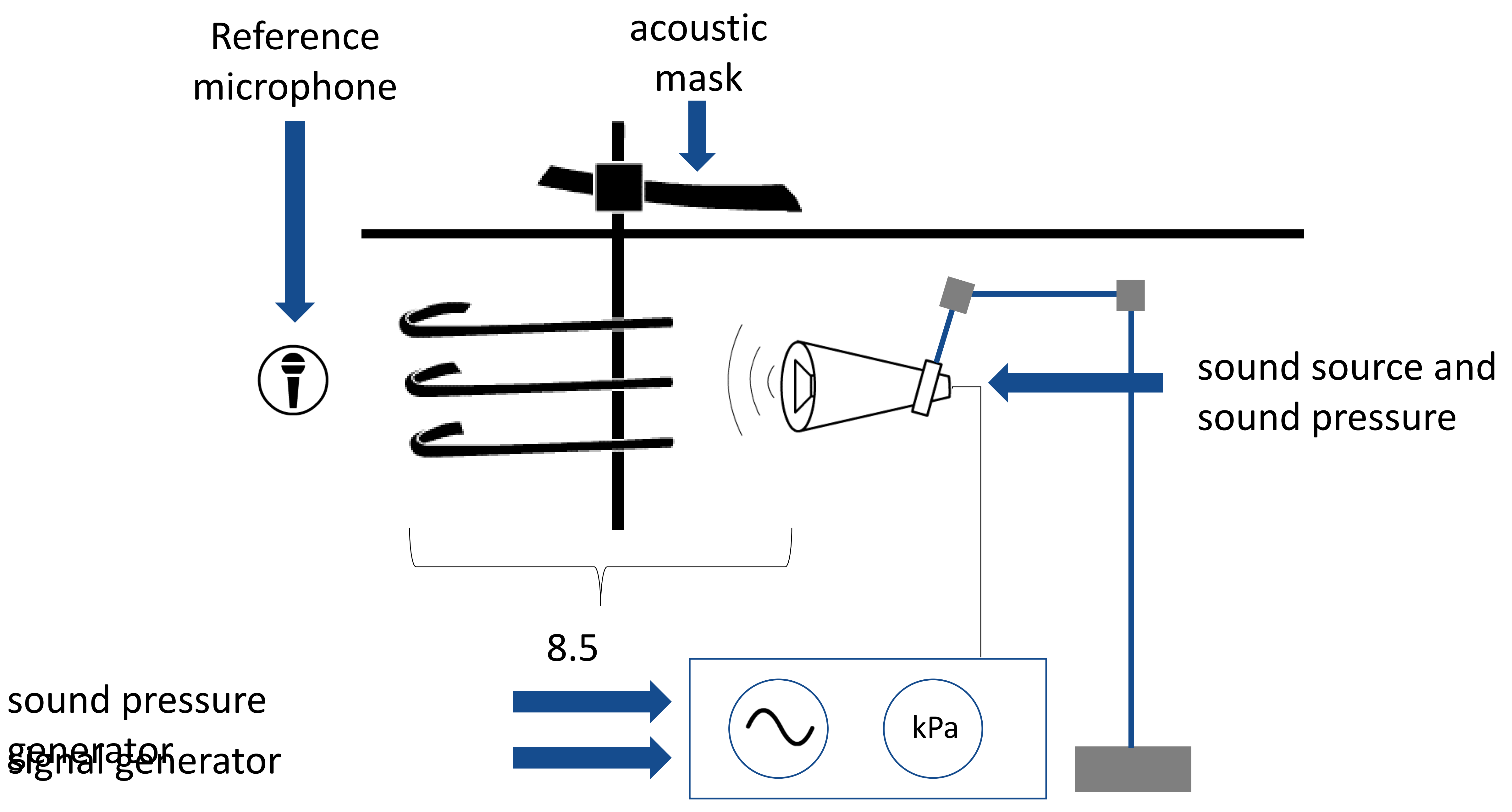
Measuring station



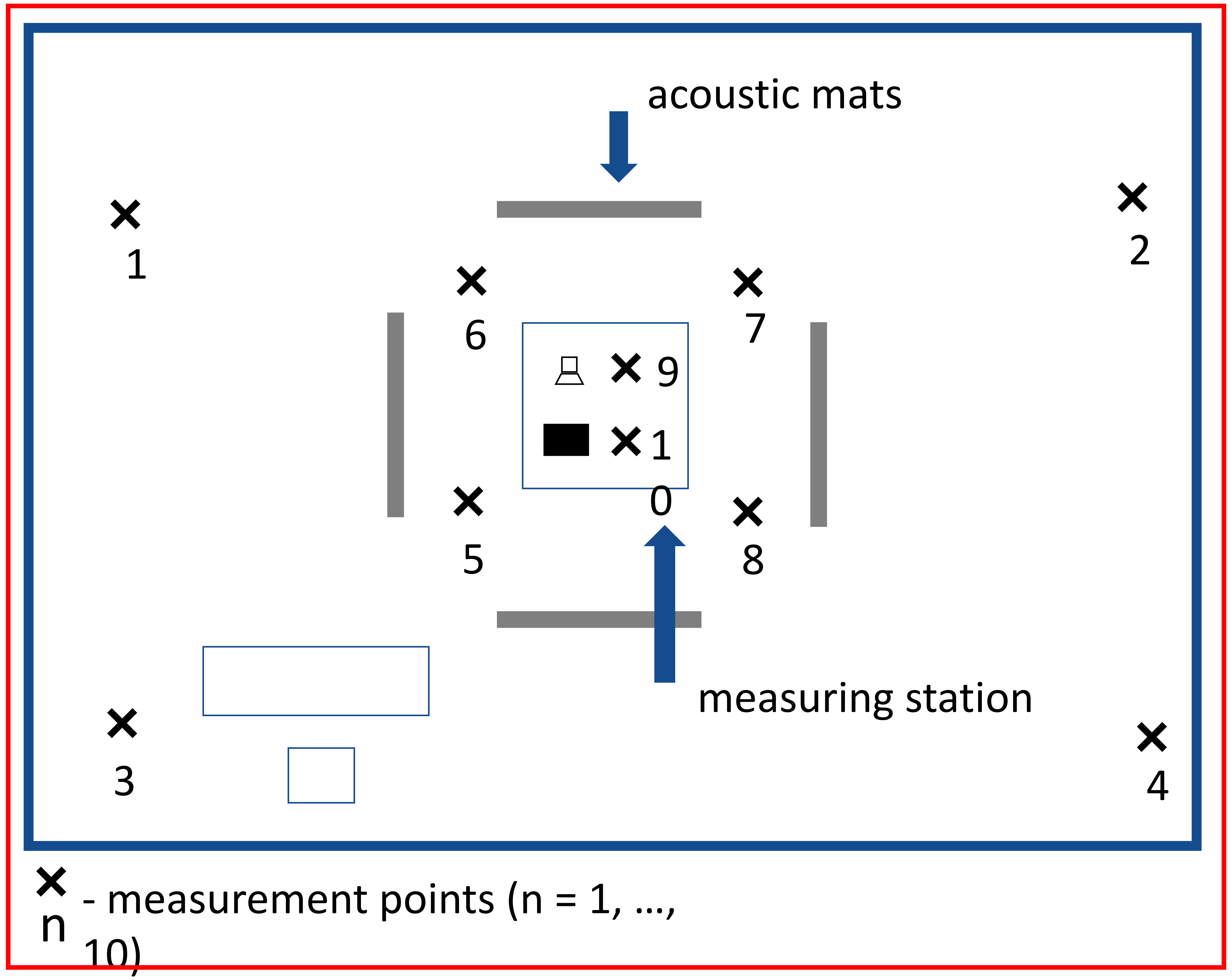
Measuring station

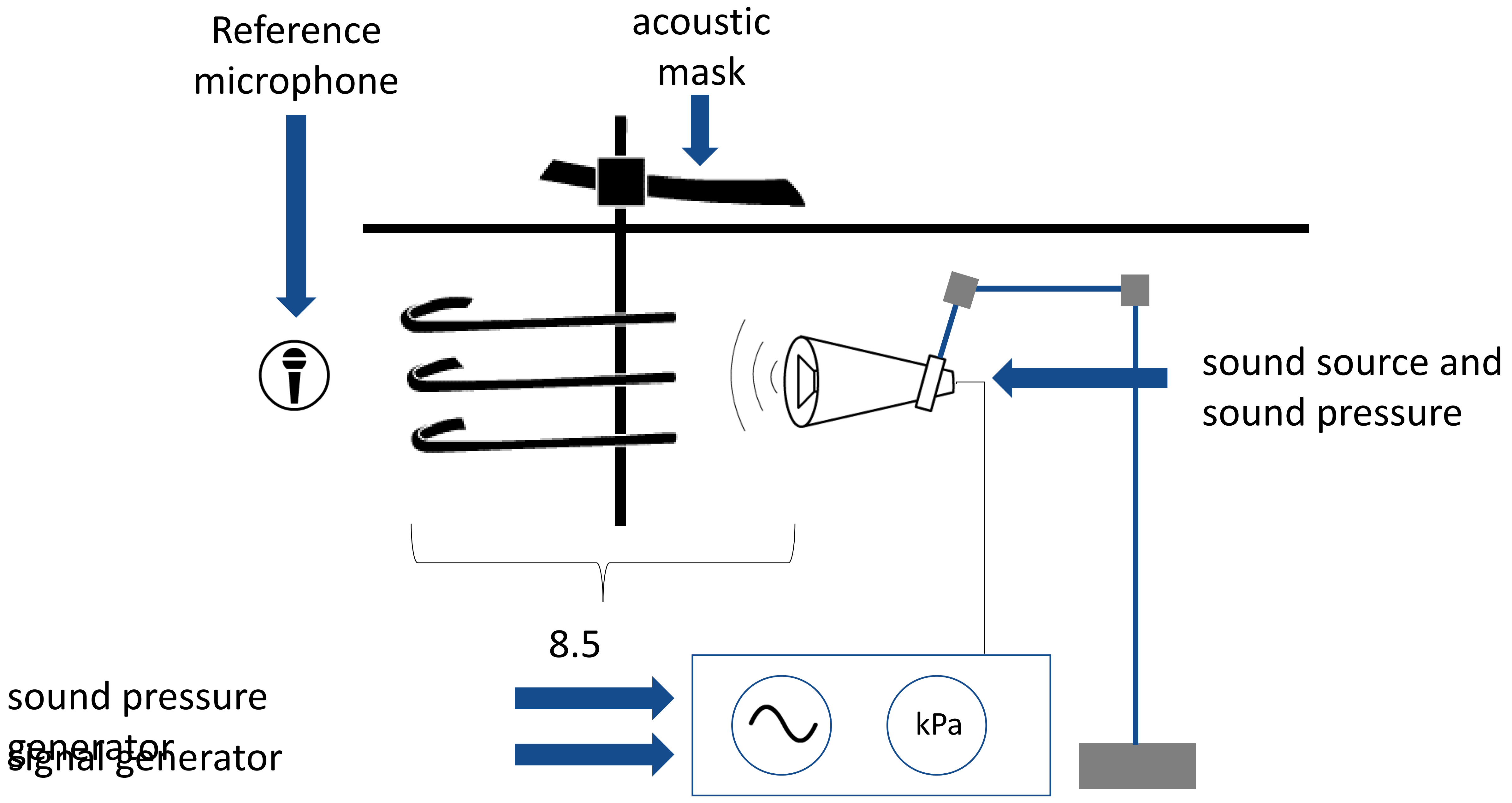


- measurement points (n = 1, …, 10)

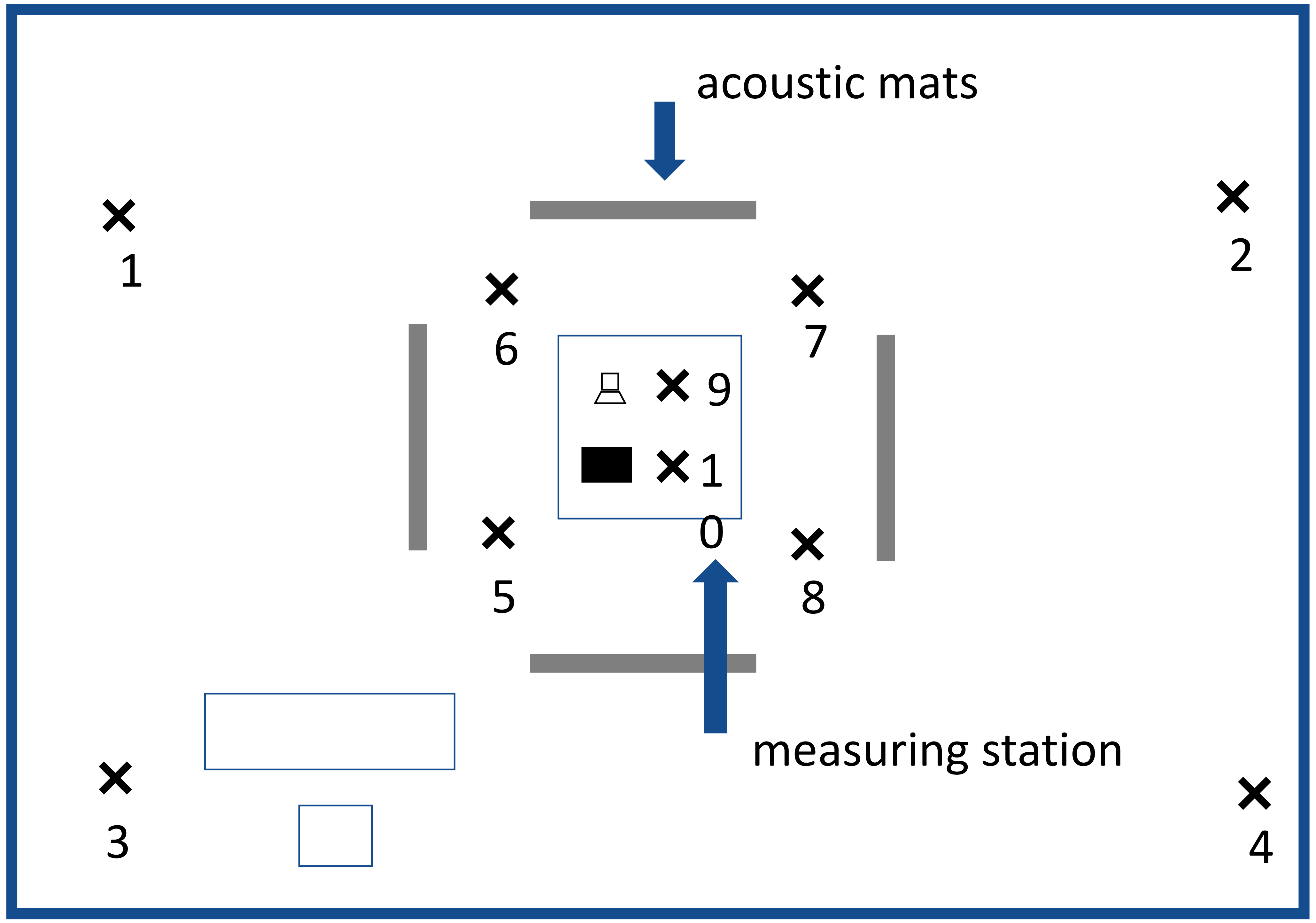


Measuring station

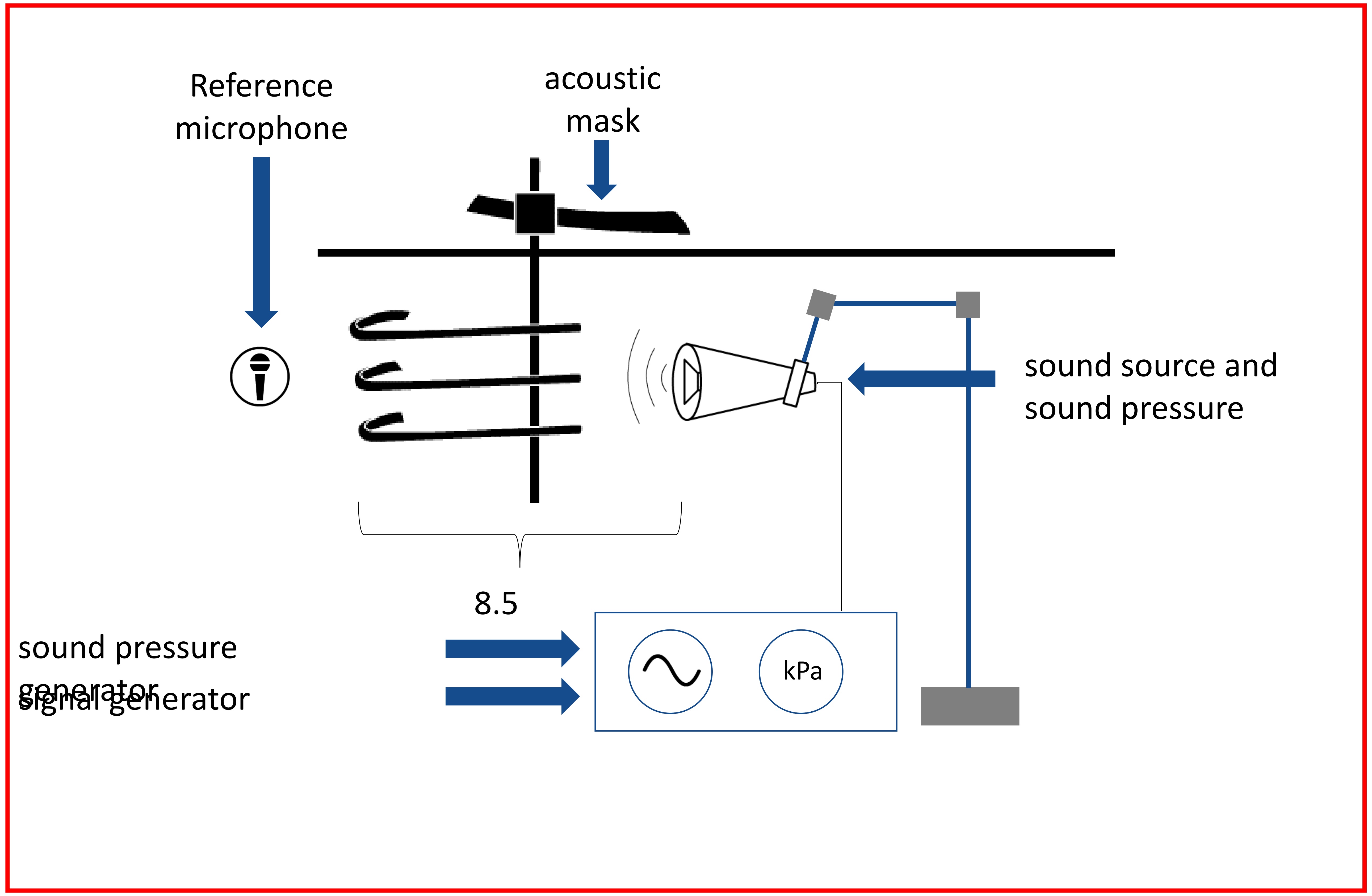




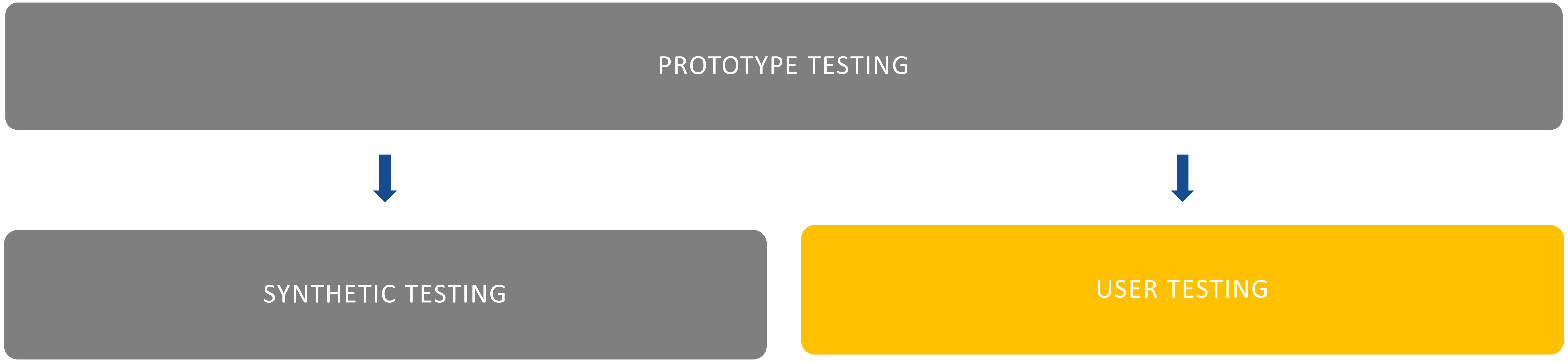
Measuring station



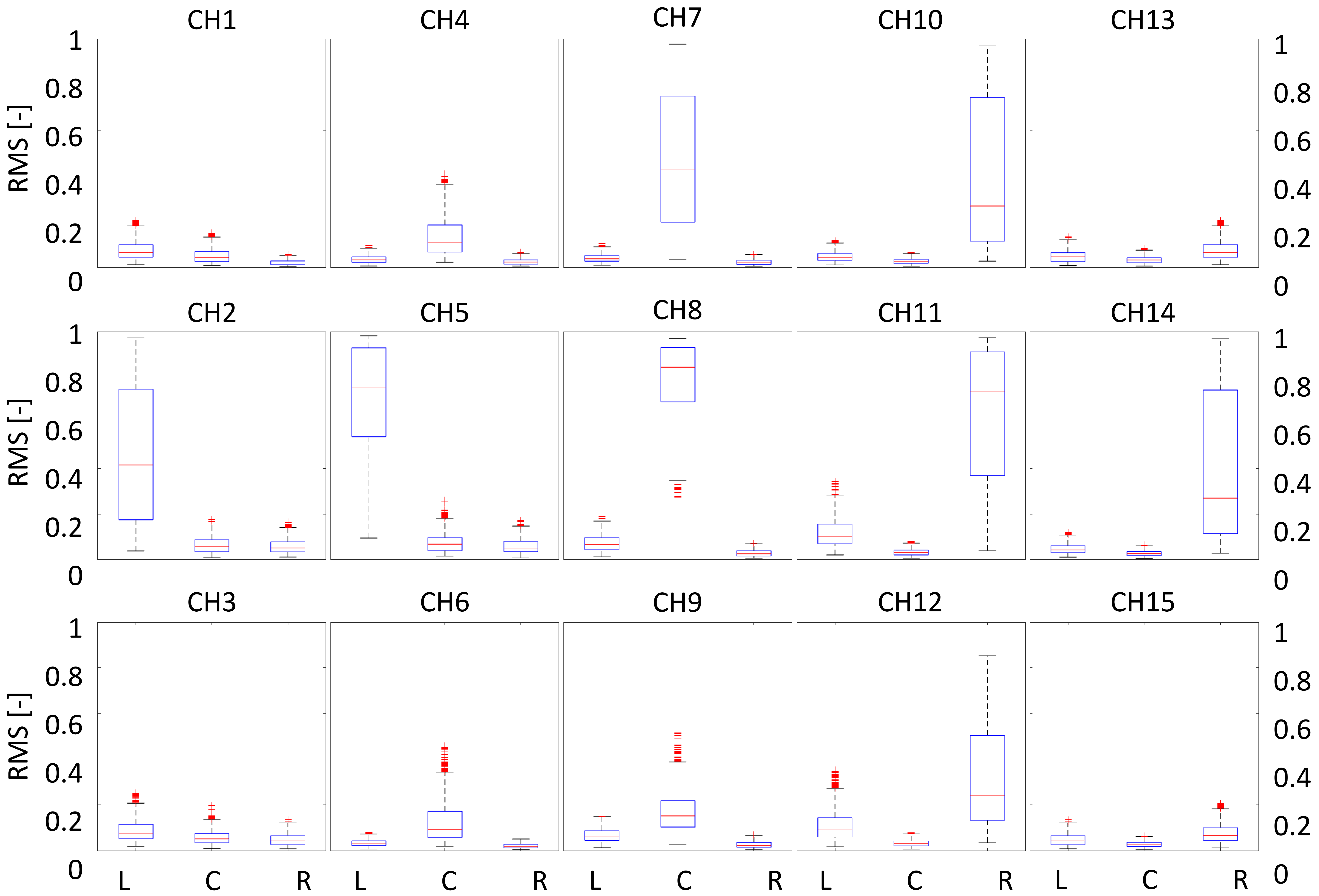
× n - measurement points (n = 1, ..., 10)



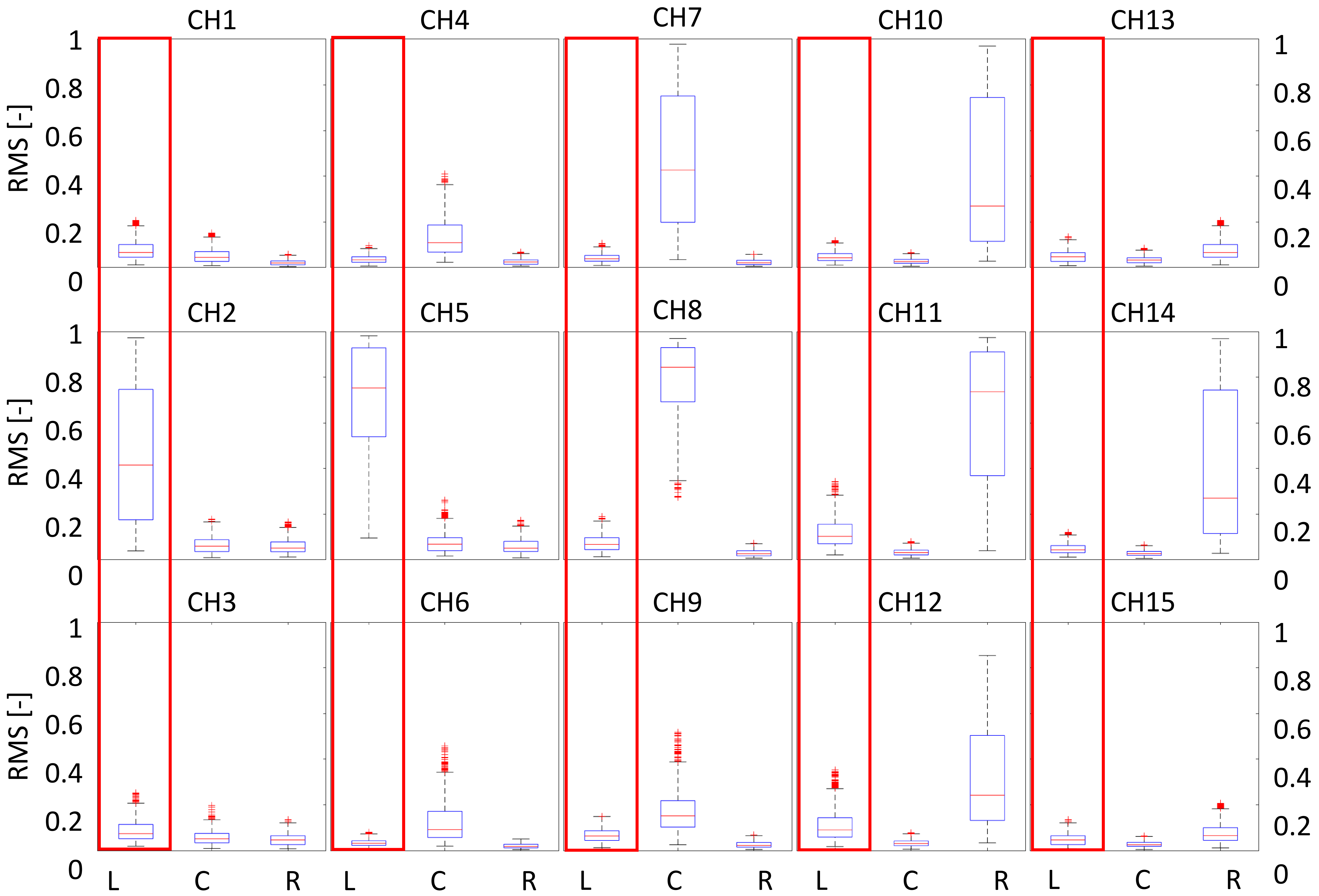
Measuring station



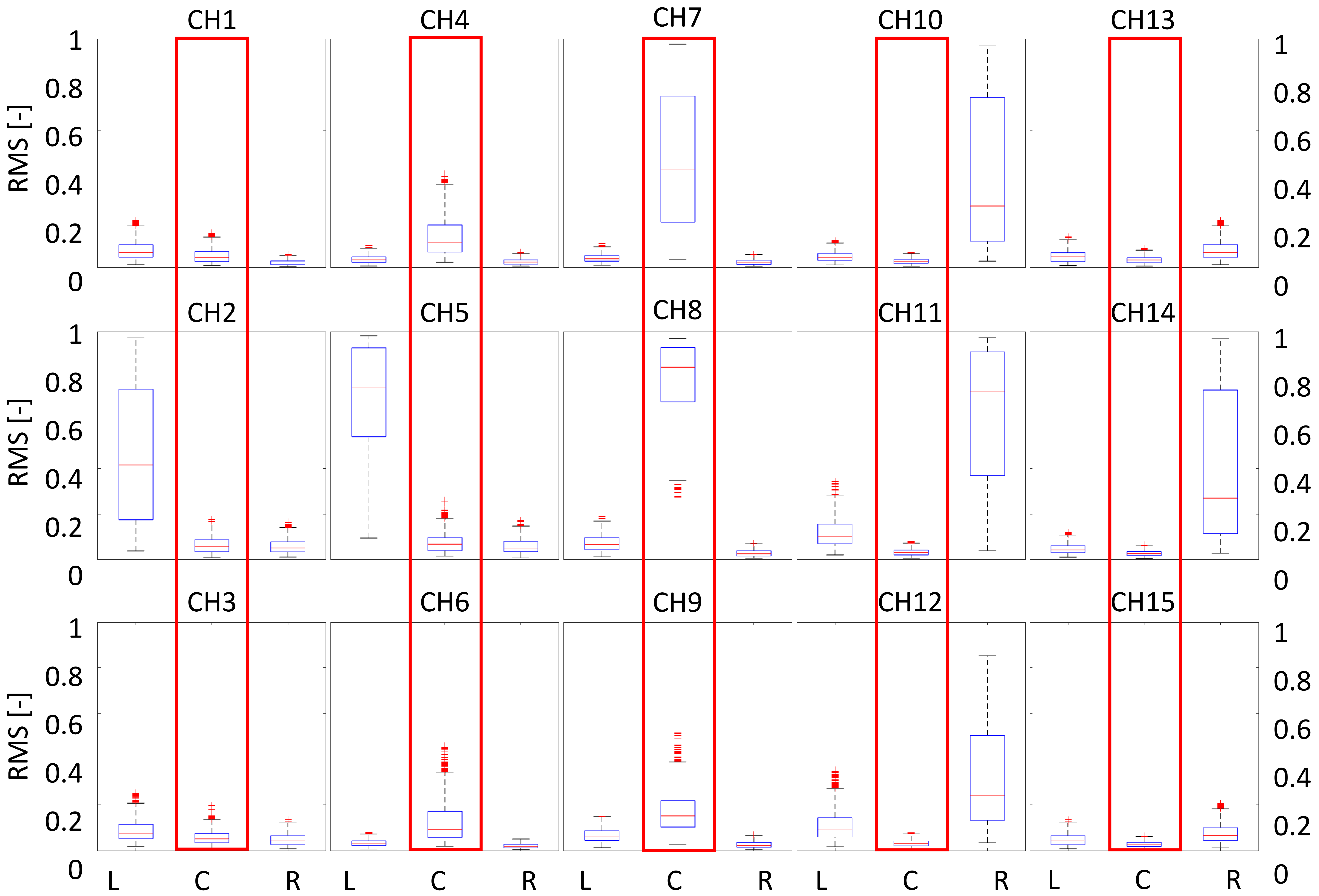
Measuring station



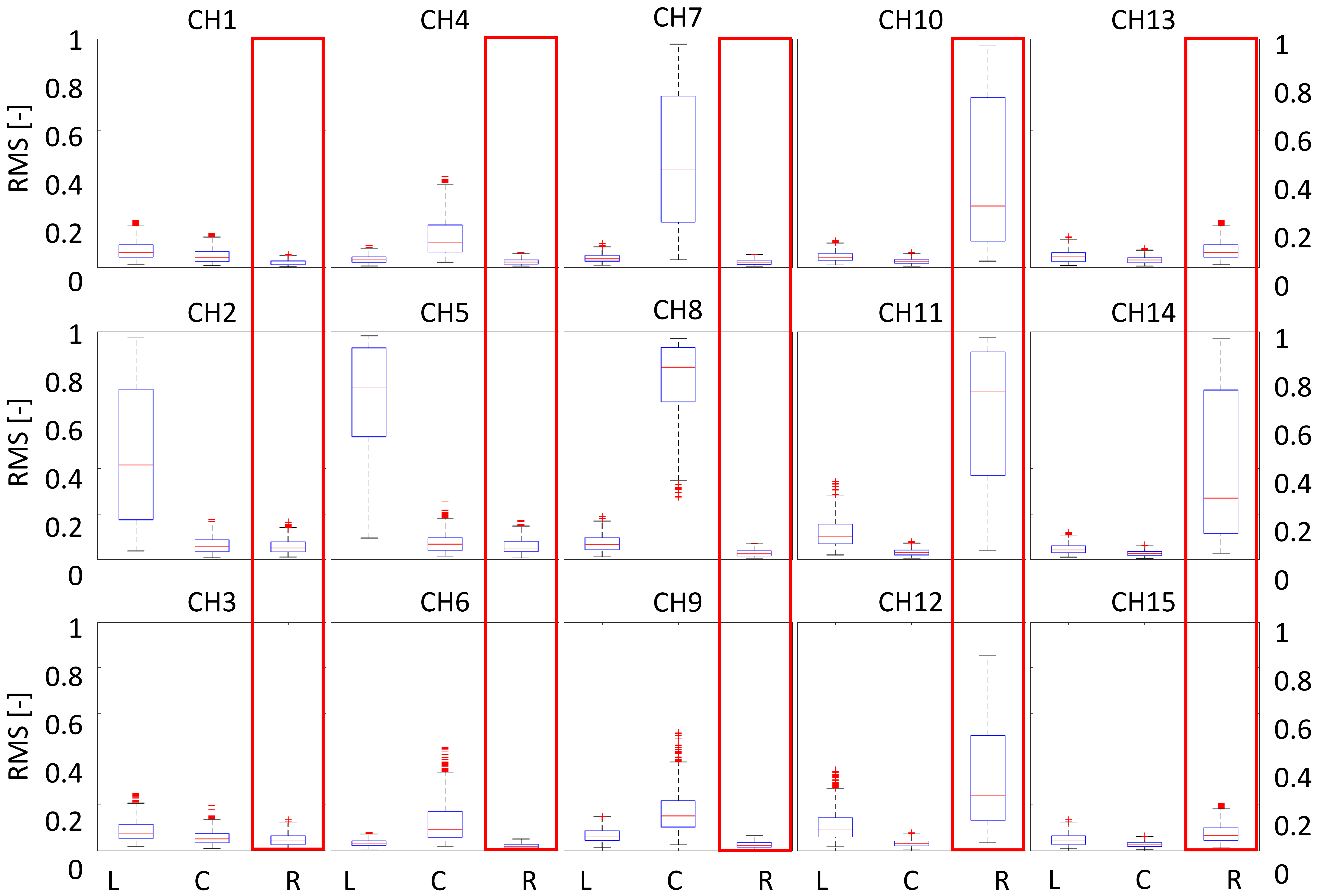
Measuring station



Measuring station



Measuring station



Measuring station

- multi-channel, spatial and repeatable sound recording,
- time synchronization,
- visual adjustment,
- size adjustment,
- semi-cylindrical surface.

Speech Database

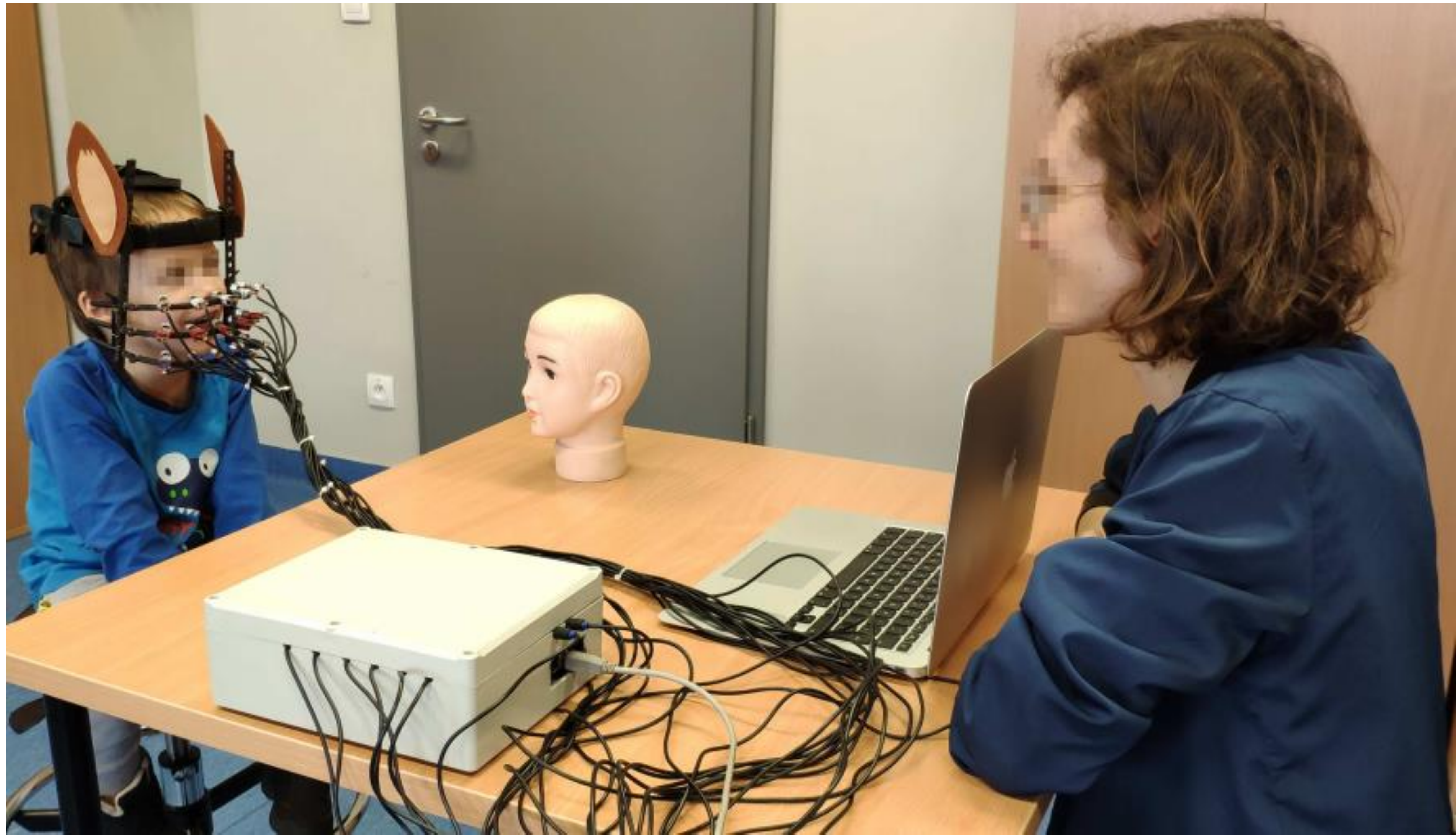
CONFIGURATION
OF THE MEASURING DEVICE

↓

PICTURE
TEST

↓

SPEECH THERAPY
DESCRIPTION



Speech Database

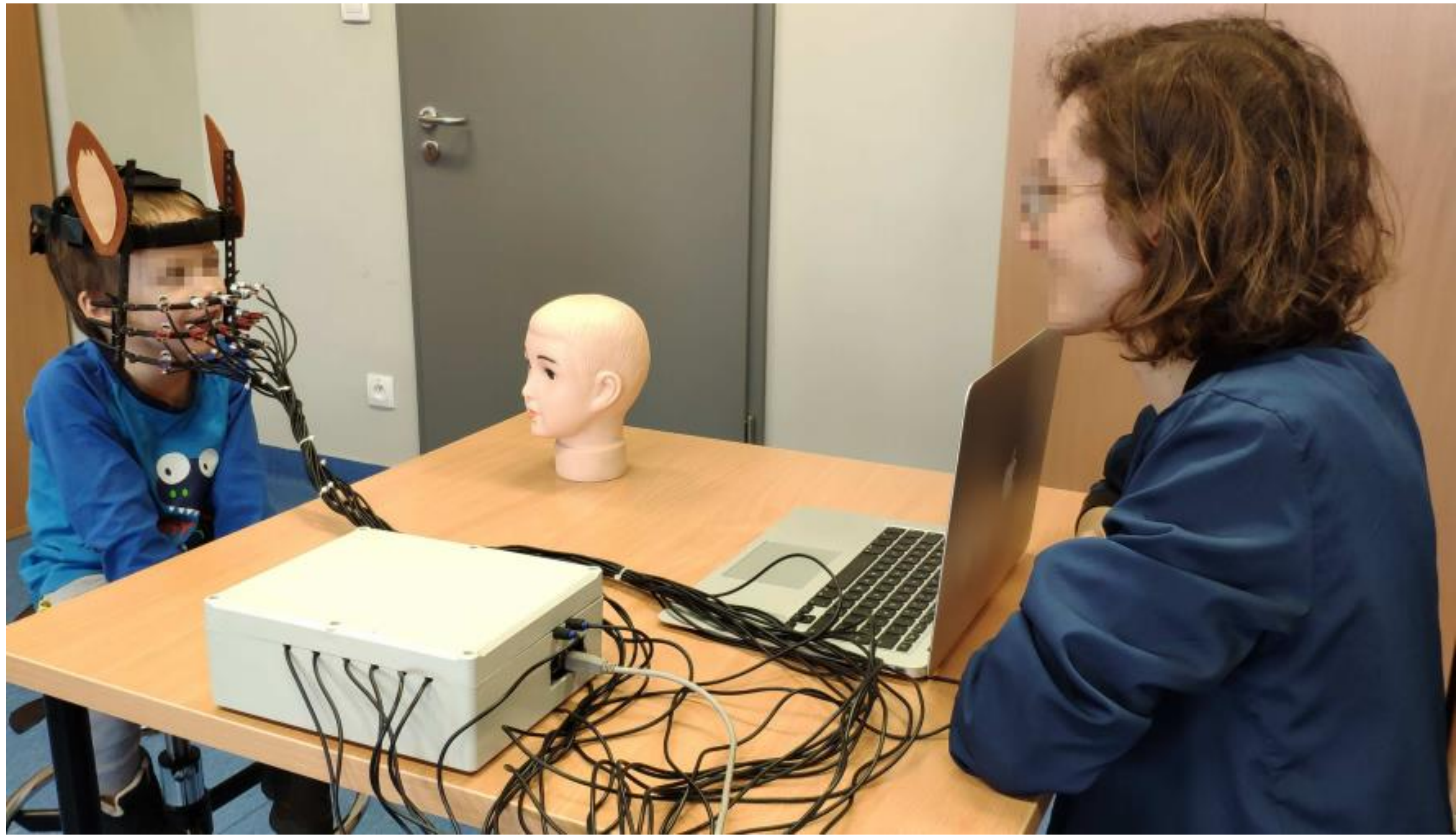
CONFIGURATION
OF THE MEASURING DEVICE

↓

PICTURE
TEST

↓

SPEECH THERAPY
DESCRIPTION



Speech Database

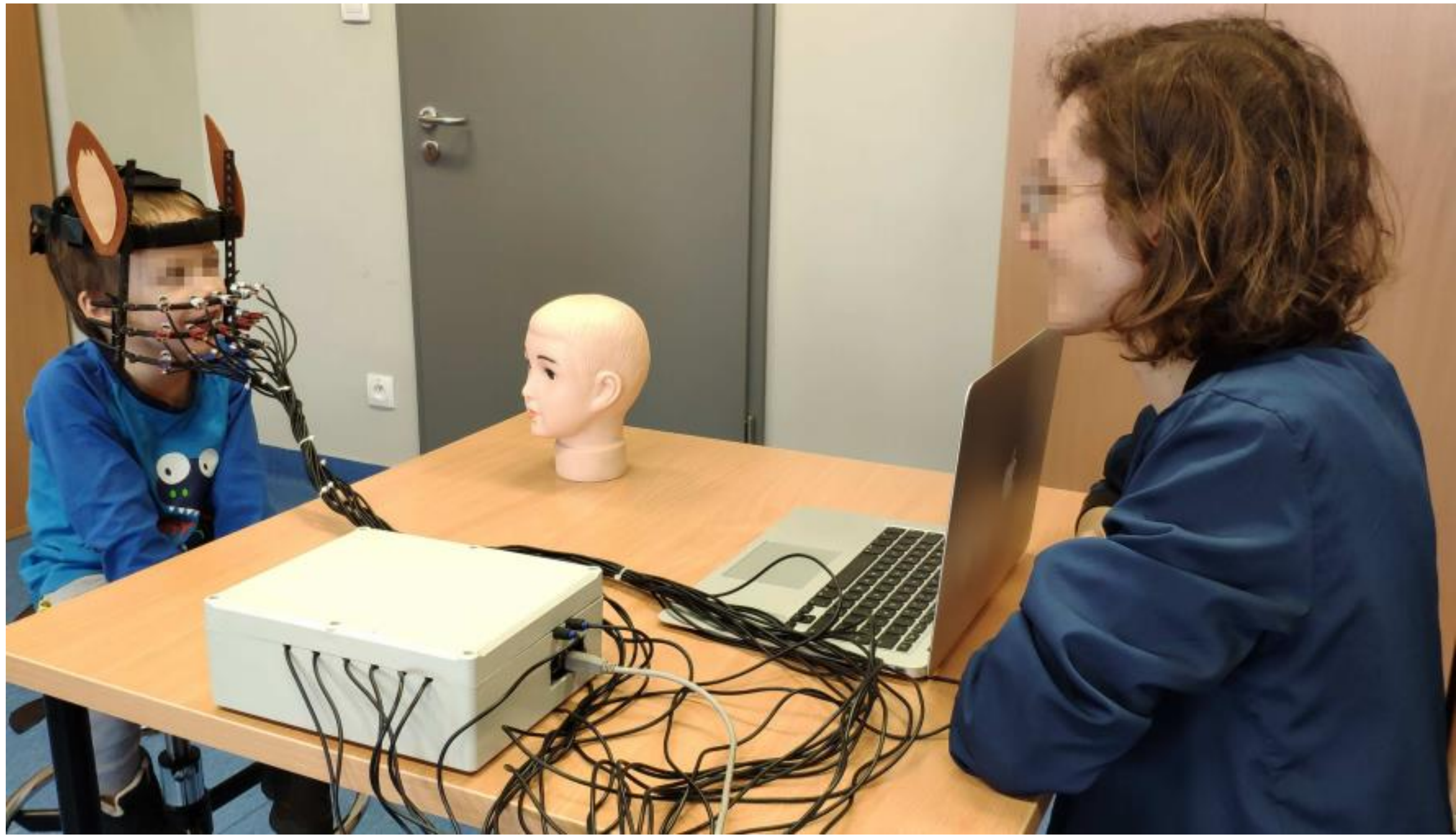
CONFIGURATION
OF THE MEASURING DEVICE

↓

PICTURE
TEST

↓

SPEECH THERAPY
DESCRIPTION



Speech Database

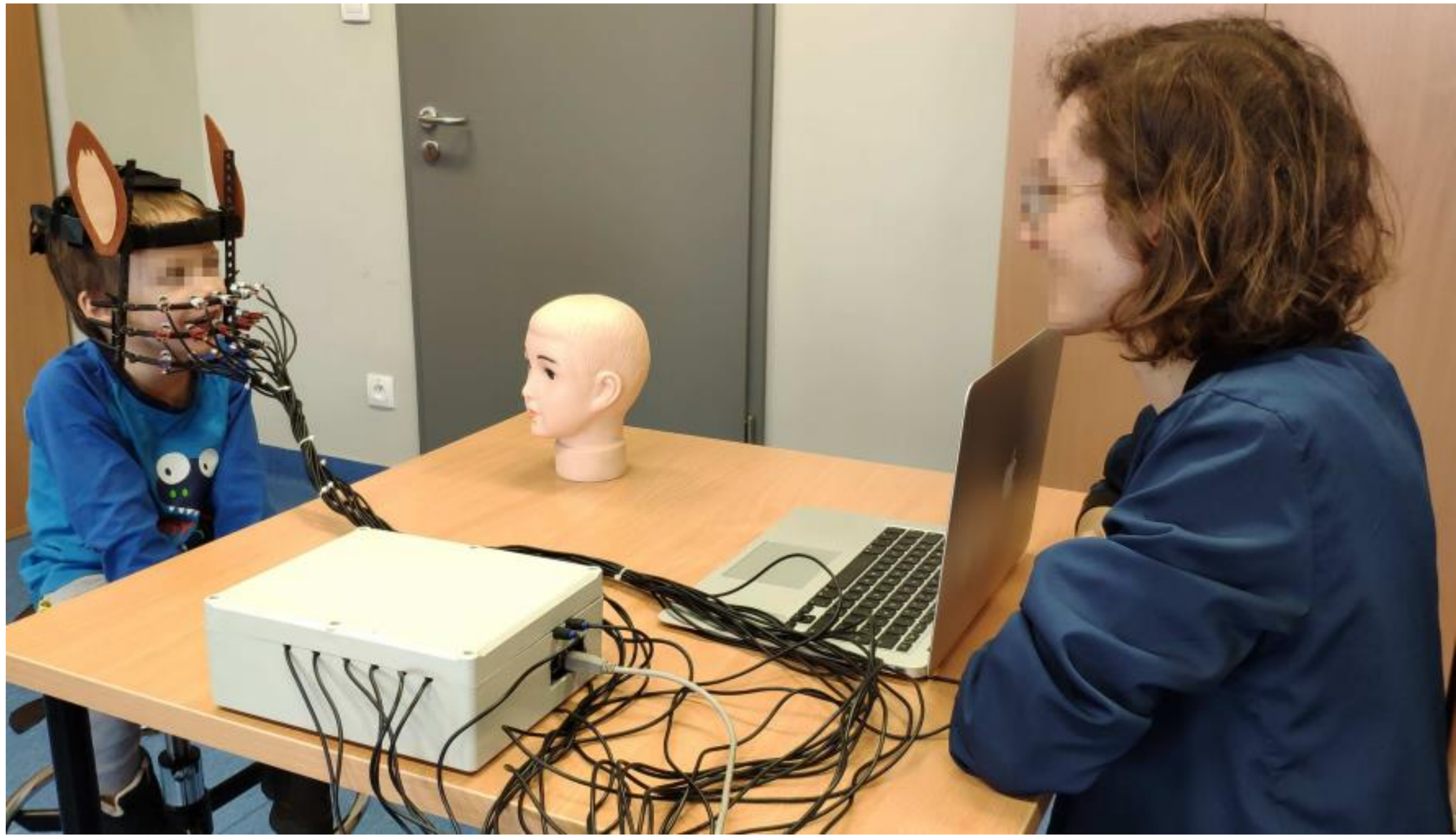
CONFIGURATION
OF THE MEASURING DEVICE

↓

PICTURE
TEST

↓

SPEECH THERAPY
DESCRIPTION



Speech Database

Dictionary

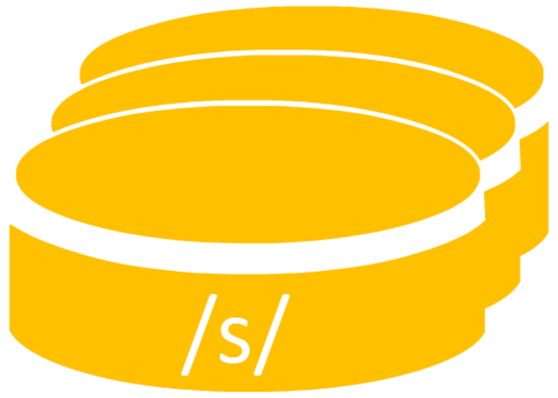
words with dentalized sounds in different articulation positions

98 speakers

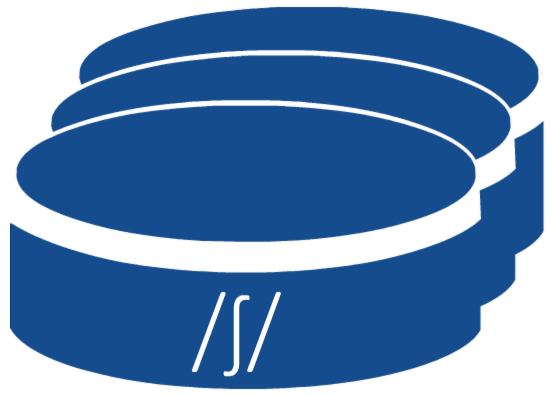
Age 6-7, 1362 recordings

Speech therapy description

normative, interdental, addental and dental realization

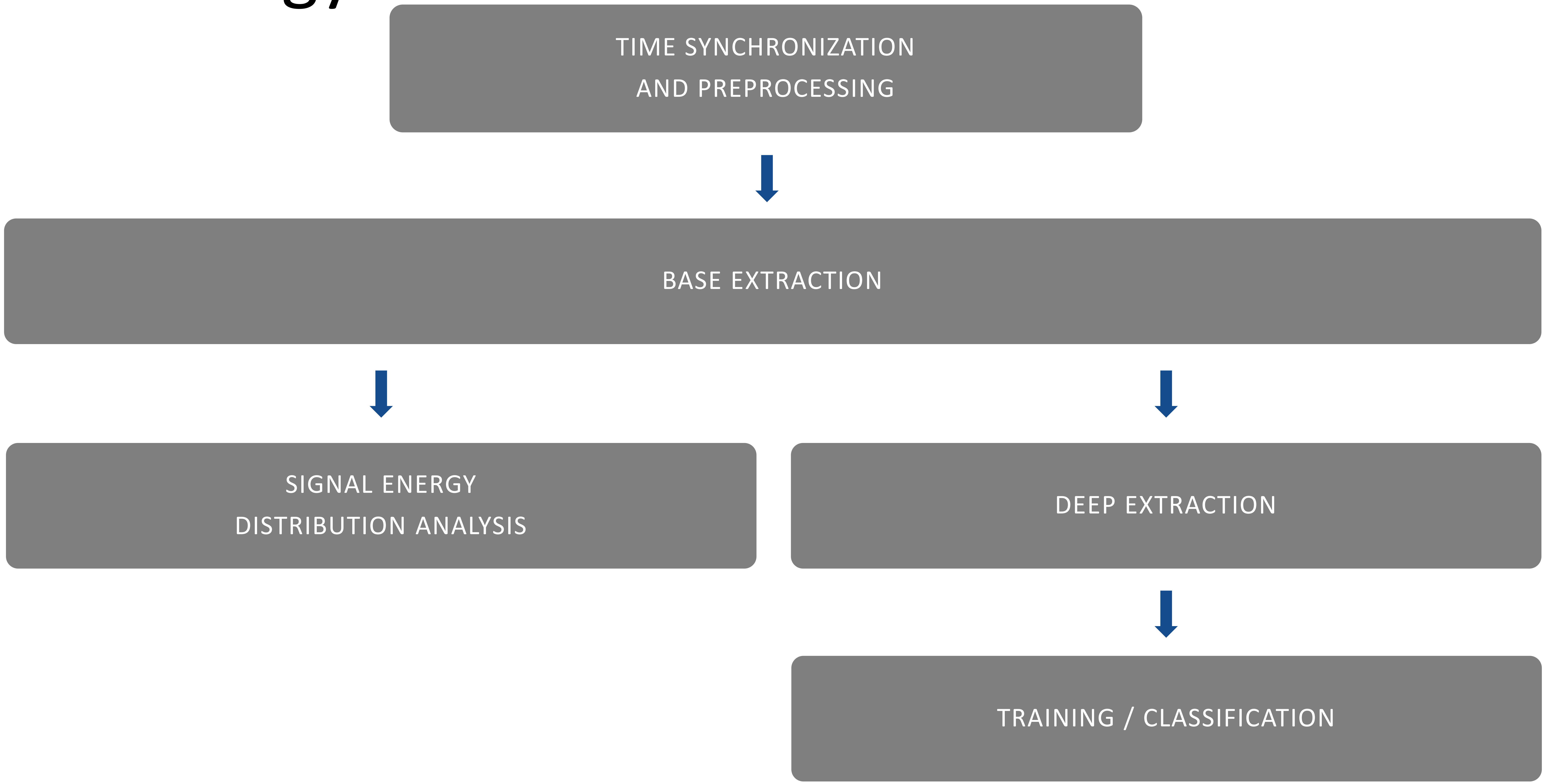


Polish	English	number
samolot	plane	76
serce	heart	82
strażak	fireman	70
pasek	belt	81
parasol	umbrella	48
pies	dog	51
lis	fox	85



Polish	English	number
szafa	wardrobe	82
sznur	rope	11
szufelka	dustpan	79
kalosze	wellies	82
koszyk	basket	82
książka	book	11
lekarz	doctor	12
nóż	knife	78
wąż	snake	82

Methodology



Methodology

TIME SYNCHRONIZATION AND PREPROCESSING

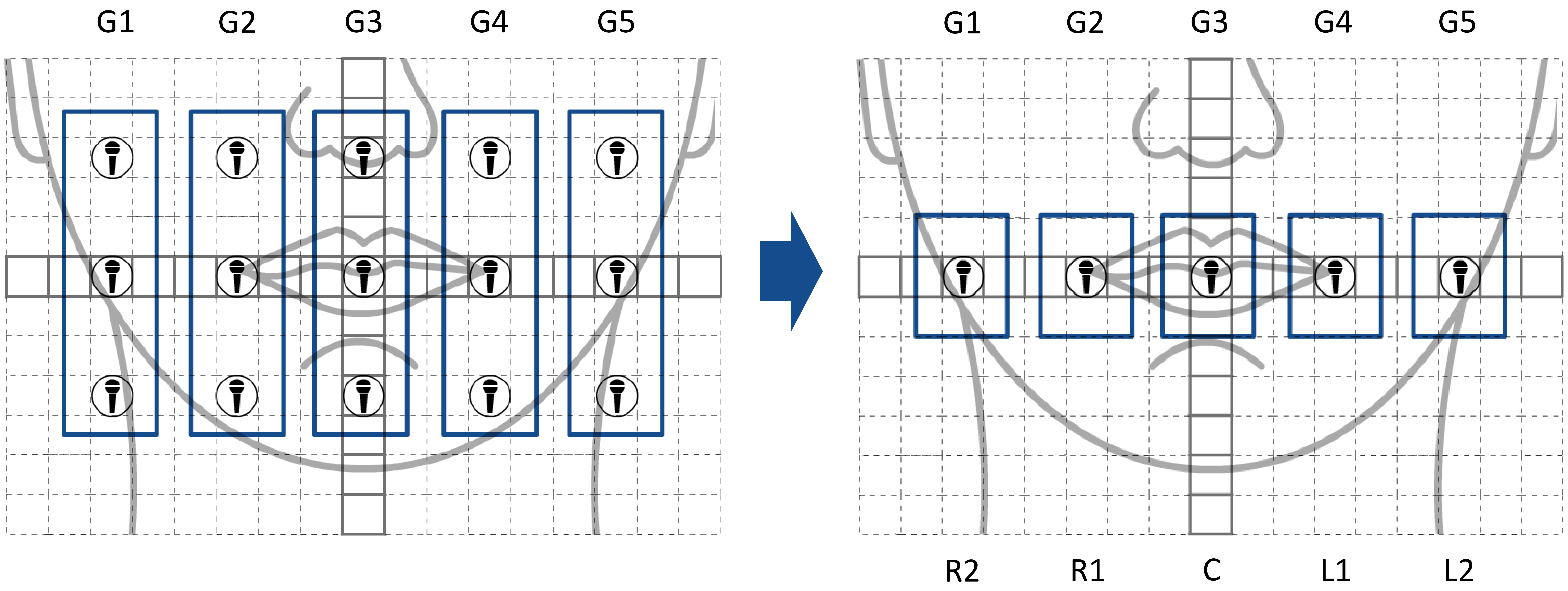
BASE EXTRACTION

SIGNAL ENERGY DISTRIBUTION ANALYSIS

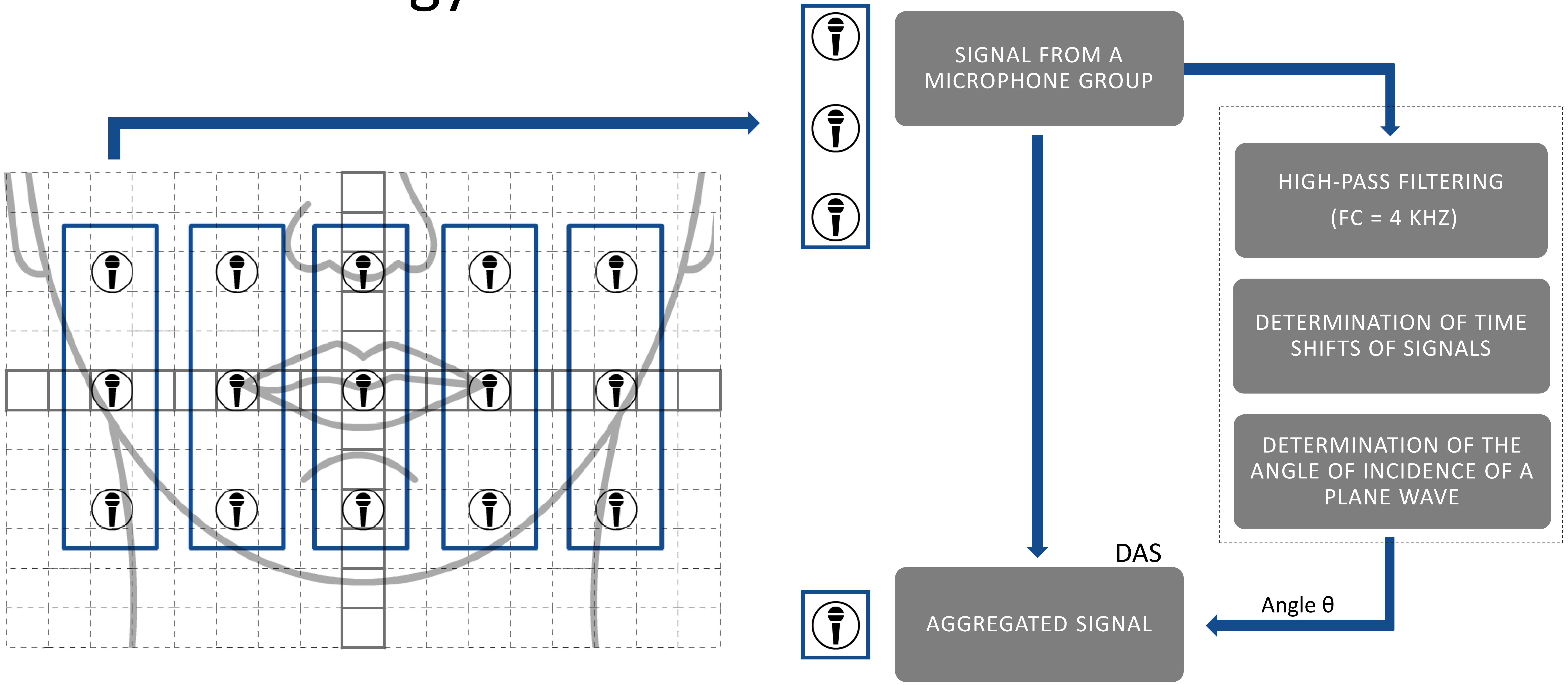
DEEP EXTRACTION

TRAINING / CLASSIFICATION

Methodology

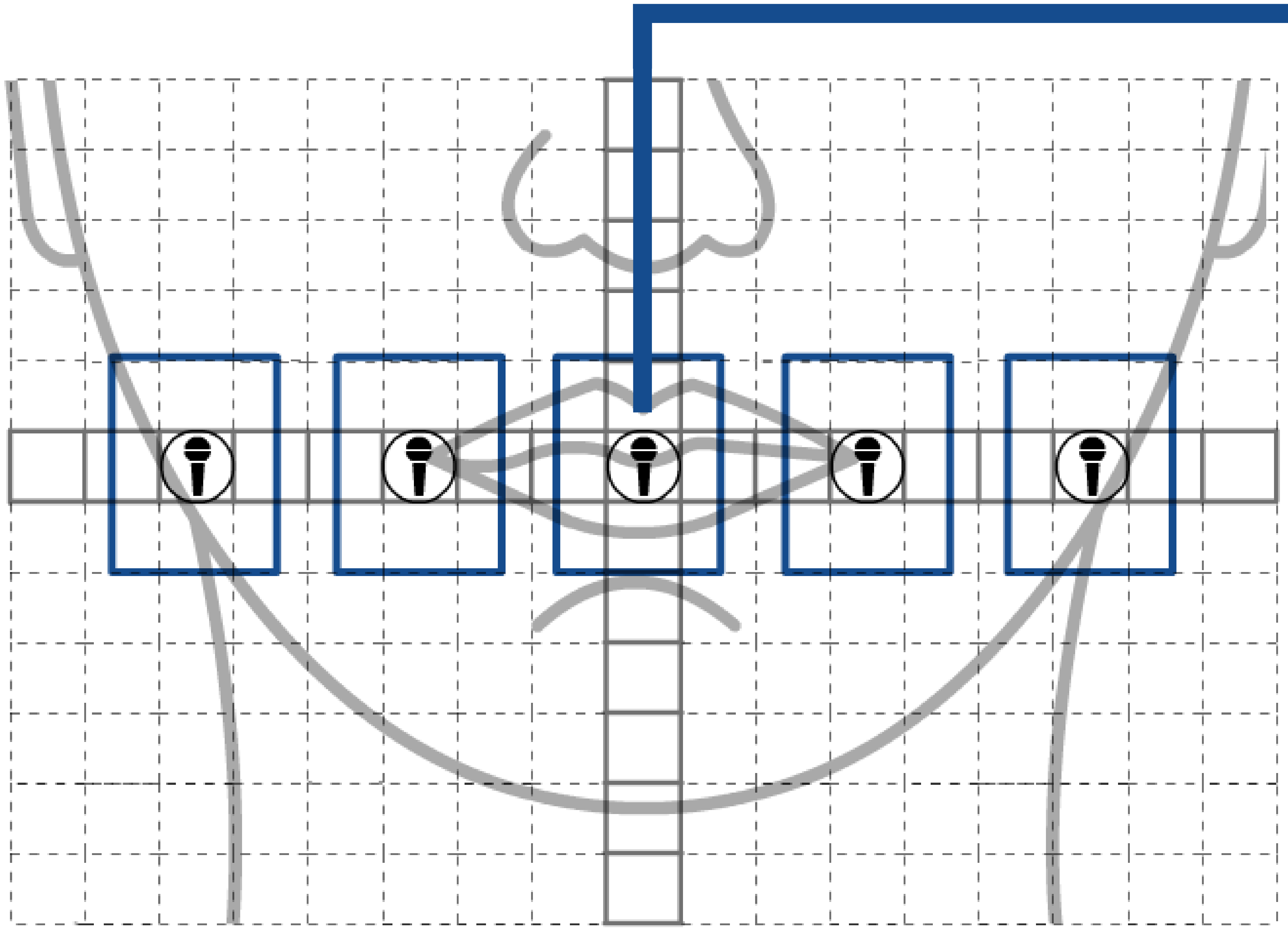


Methodology



P R E - P R O C E S S I N G

Methodology



MANUAL SEGMENTATION

FILTRACJA

PREEMPHASIS (ALPHA = 0.93)*

NORMALIZATION [-32768, 32767] - 16BIT**

FRAMING AND WINDOWING*

DURATION = 15MS

OFFSET = 5MS

HAMMING WINDOW

* - independently for each channel

** - based on all channels

Methodology

TIME SYNCHRONIZATION AND PREPROCESSING

↓

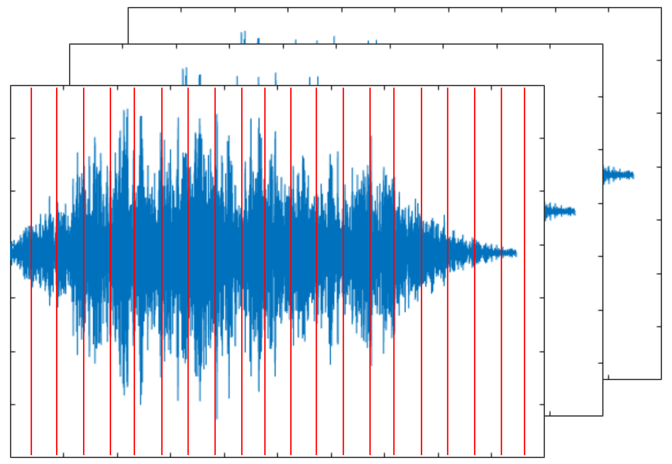
BASE EXTRACTION

↓ SIGNAL ENERGY DISTRIBUTION ANALYSIS

↓ DEEP EXTRACTION

↓ TRAINING / CLASSIFICATION

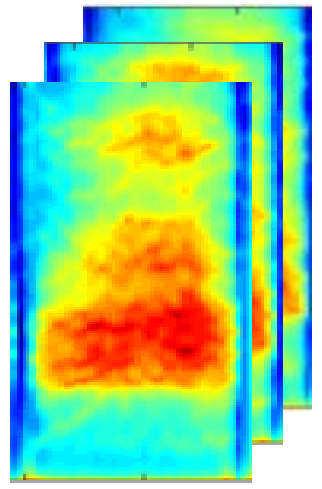
Methodology



RMS

DETERMINATION OF LPC COEFFICIENTS

PREDICTION RANK: 34

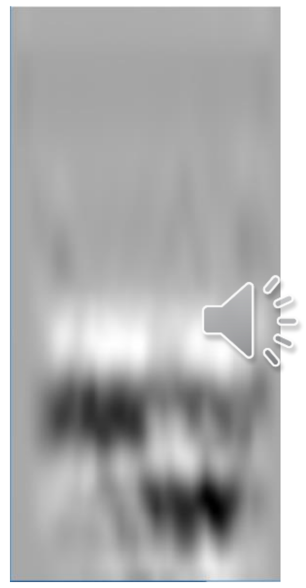


FFT

FEATURE EXTRACTION

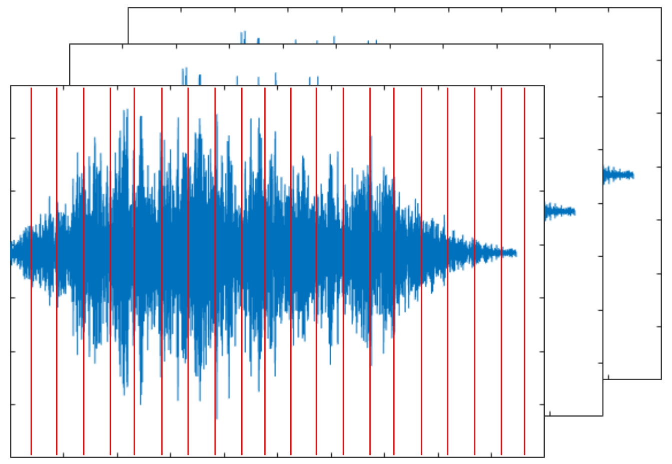
Filter Bank Energy (FBE)

Δ FBE



ΔΔ FBE

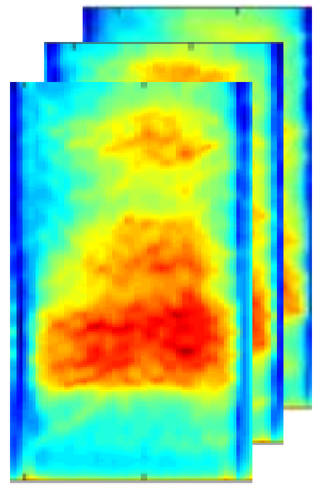
Methodology



RMS

DETERMINATION OF LPC COEFFICIENTS

PREDICTION RANK: 34

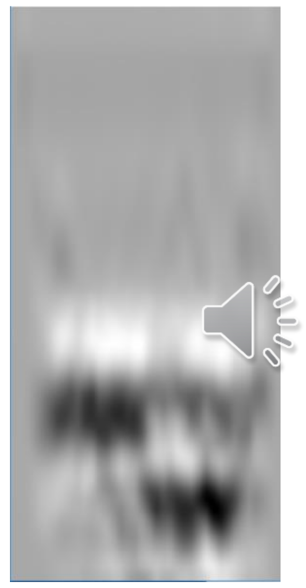


FFT

FEATURE EXTRACTION

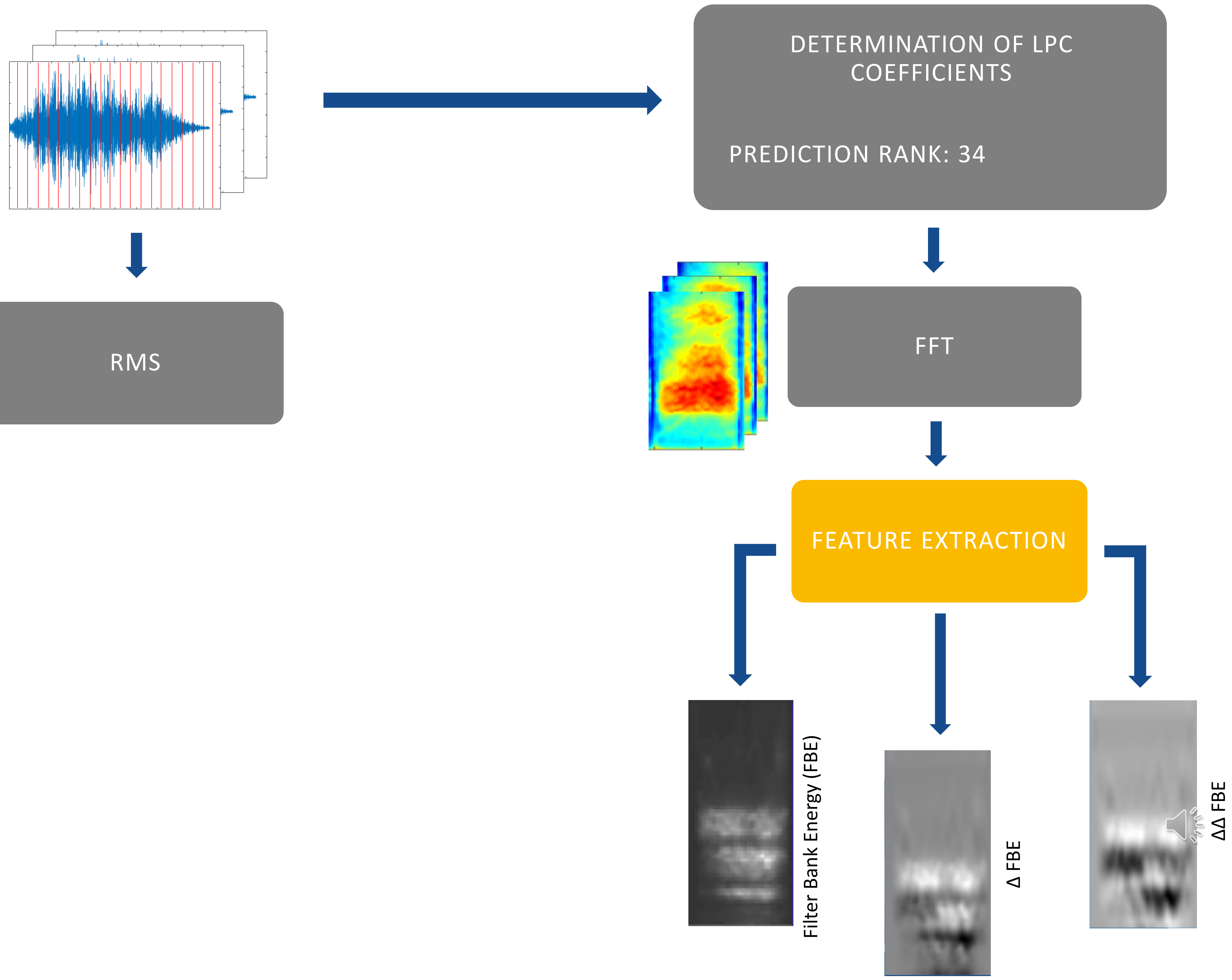
Filter Bank Energy (FBE)

Δ FBE

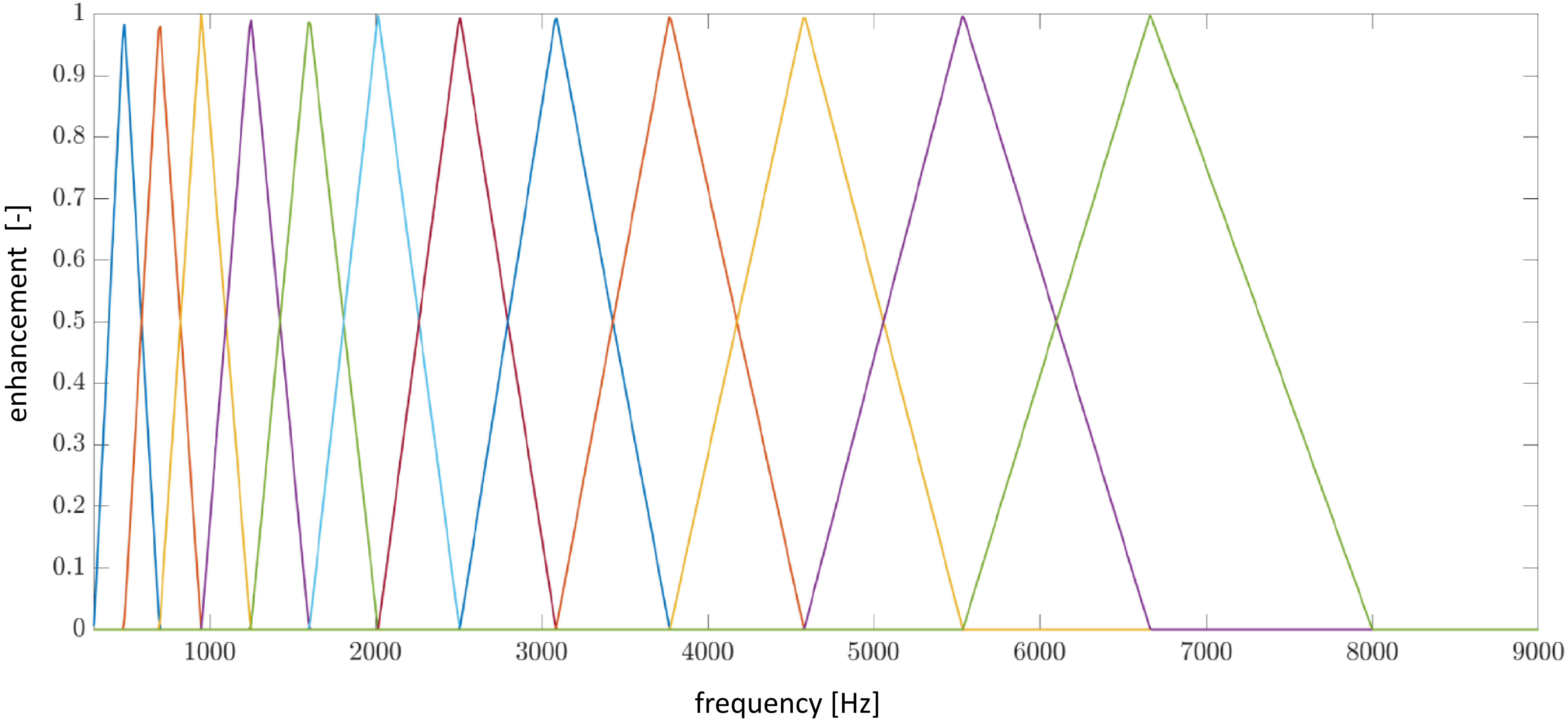


ΔΔ FBE

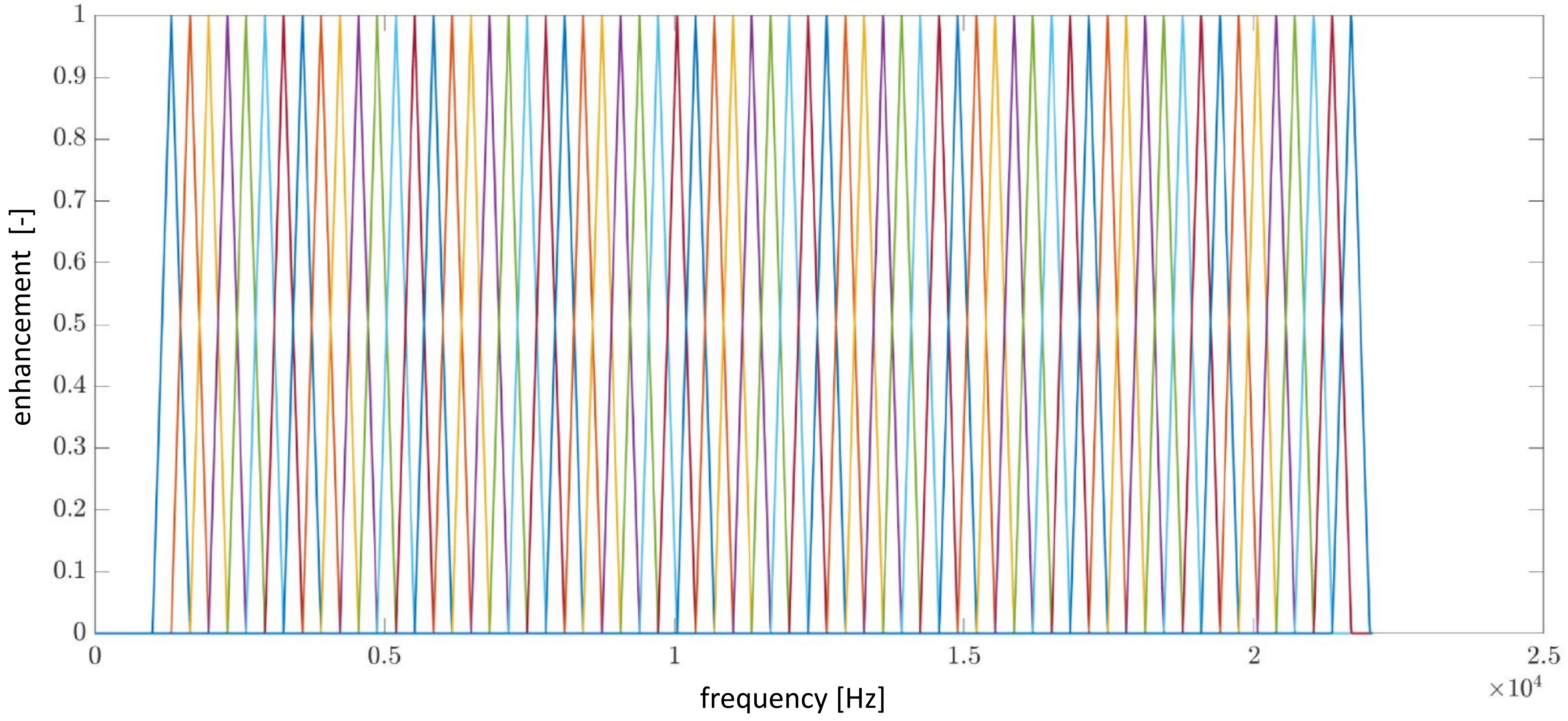
Methodology



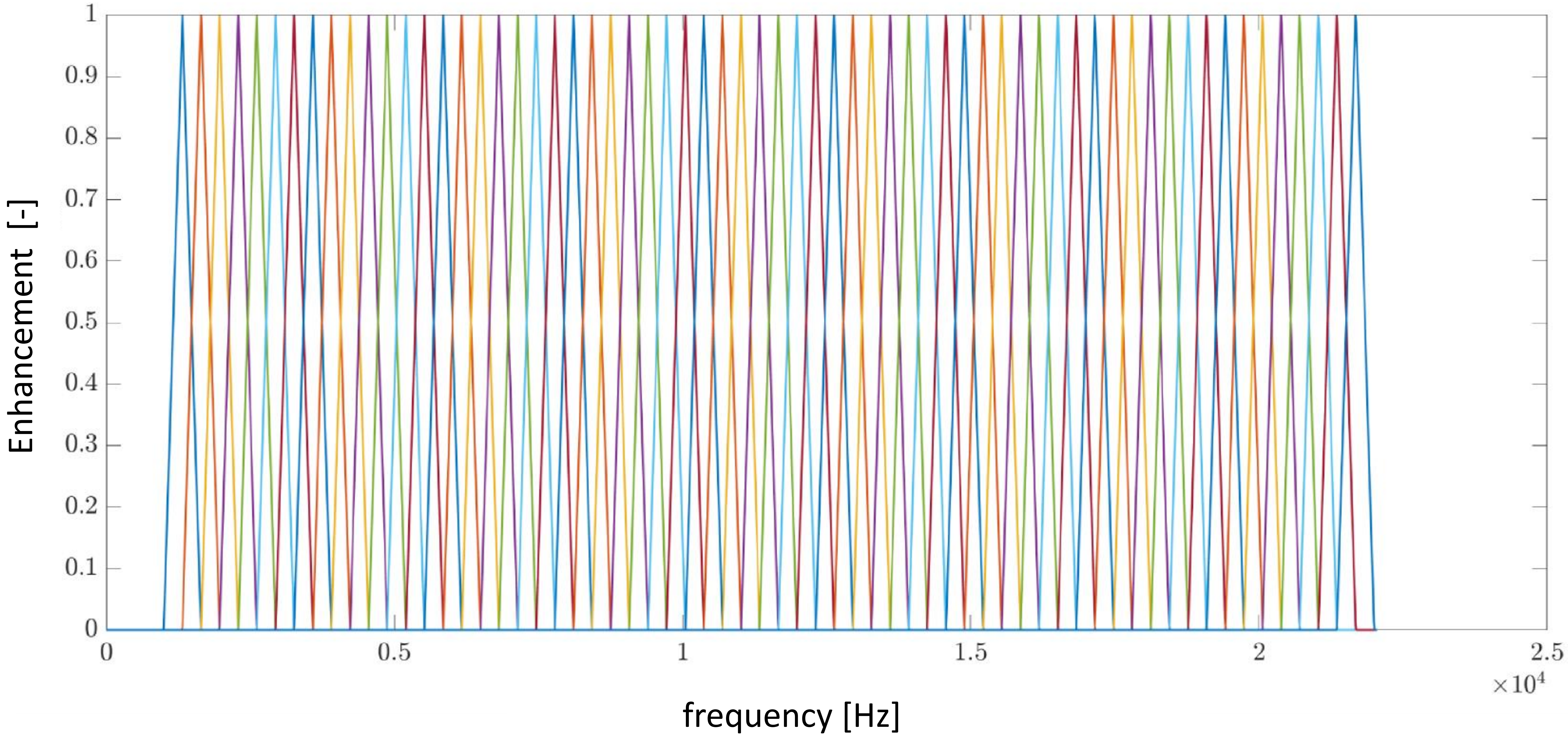
Methodology

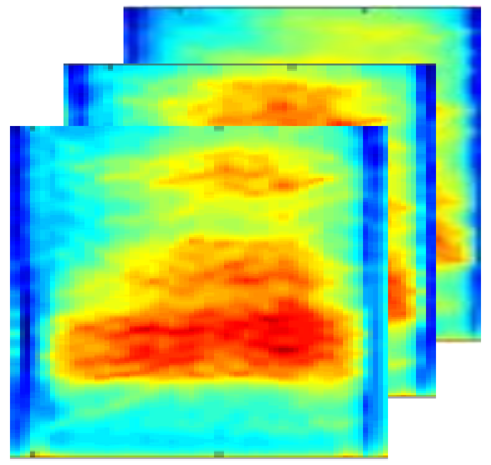


Methodology

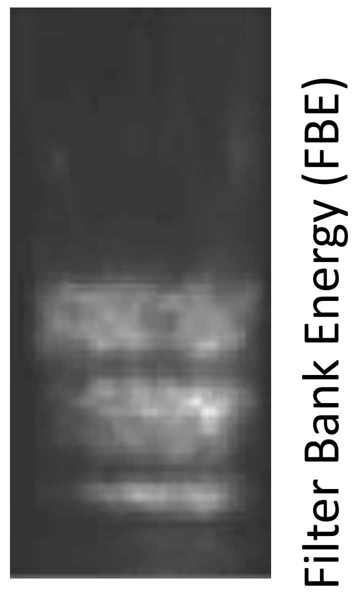


Methodology



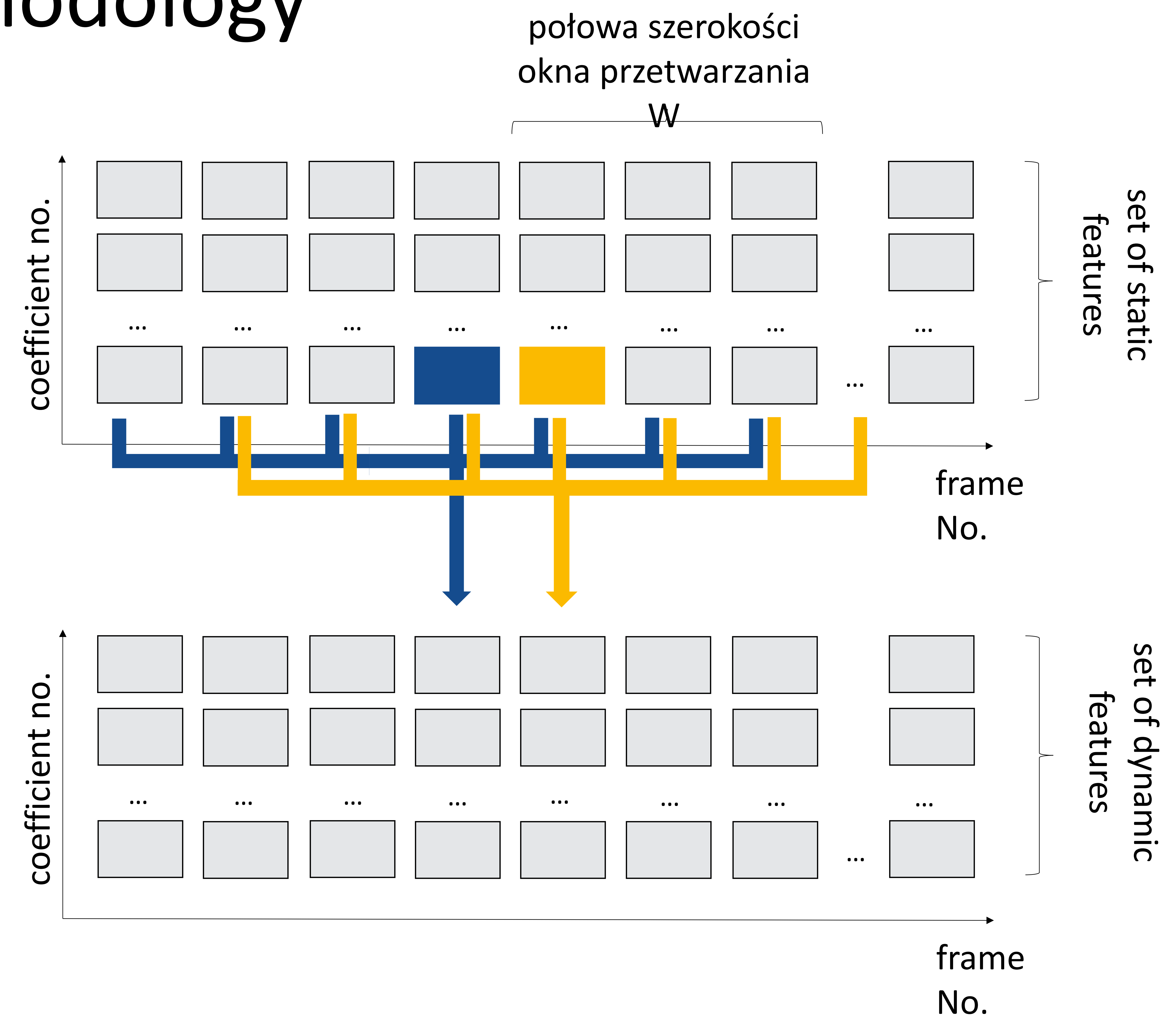


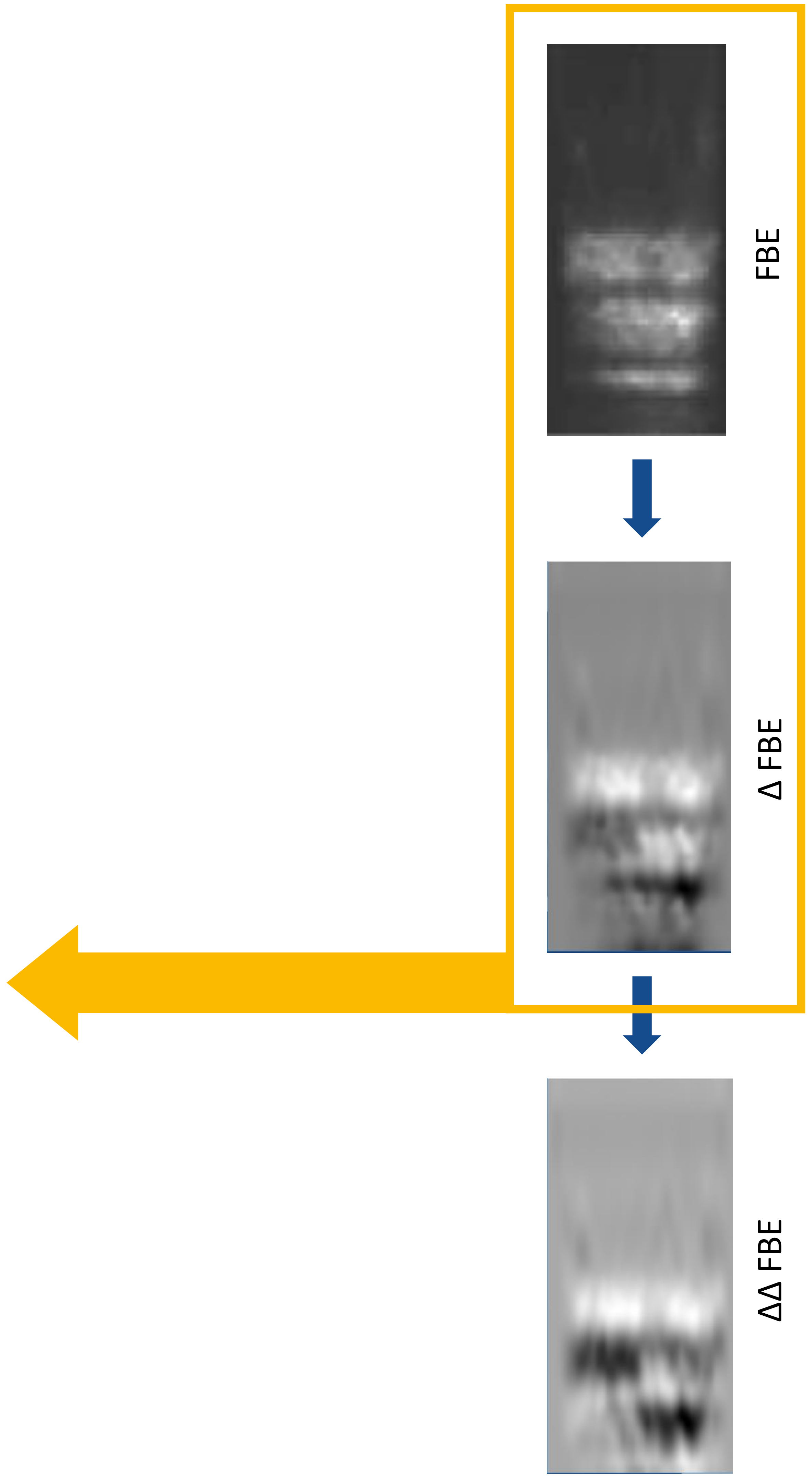
FILTERING



Methodology

np.: FBE 64 x 30



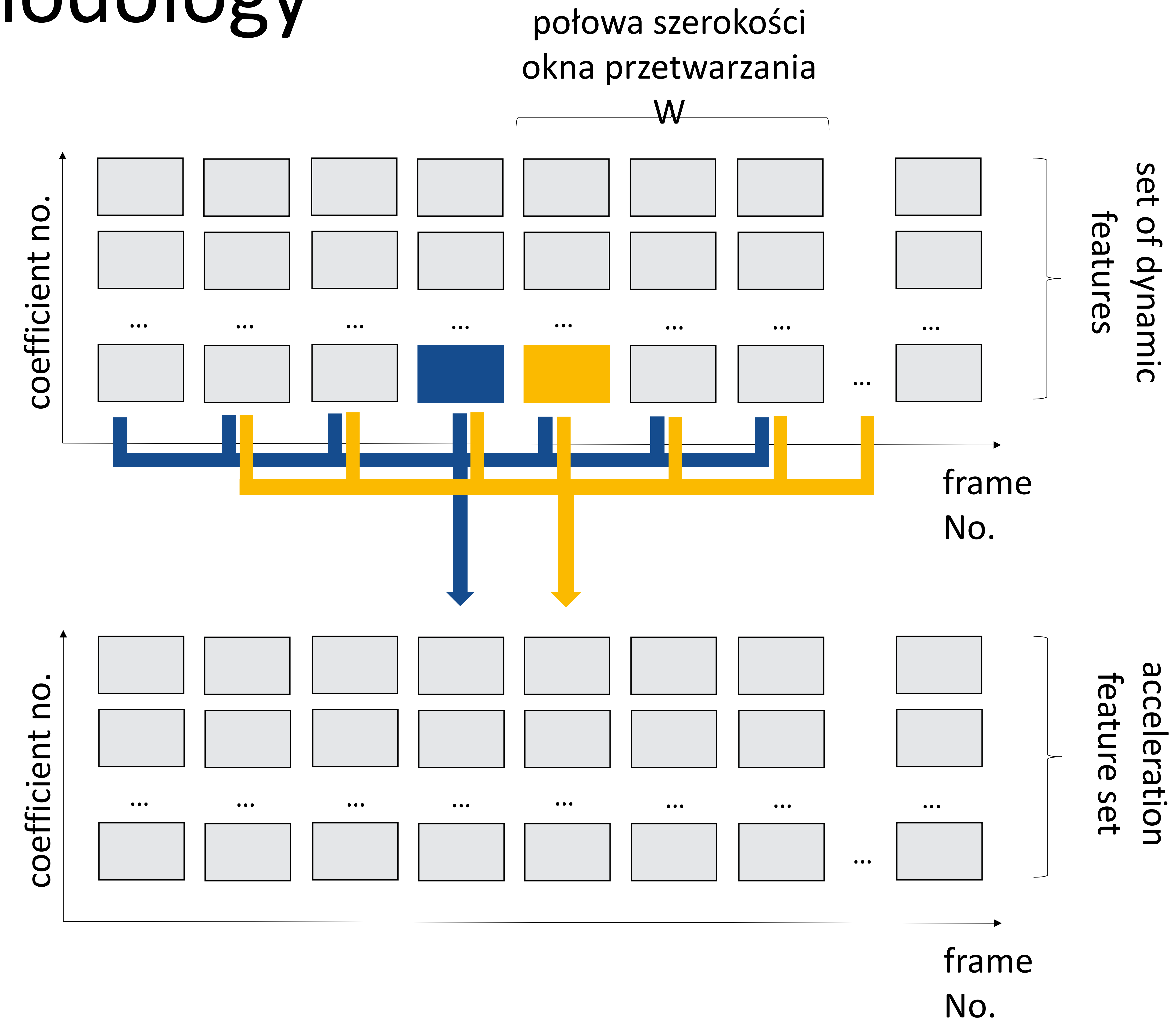


np.: FBE 64 x 30
+ ΔFBE 64 x 30

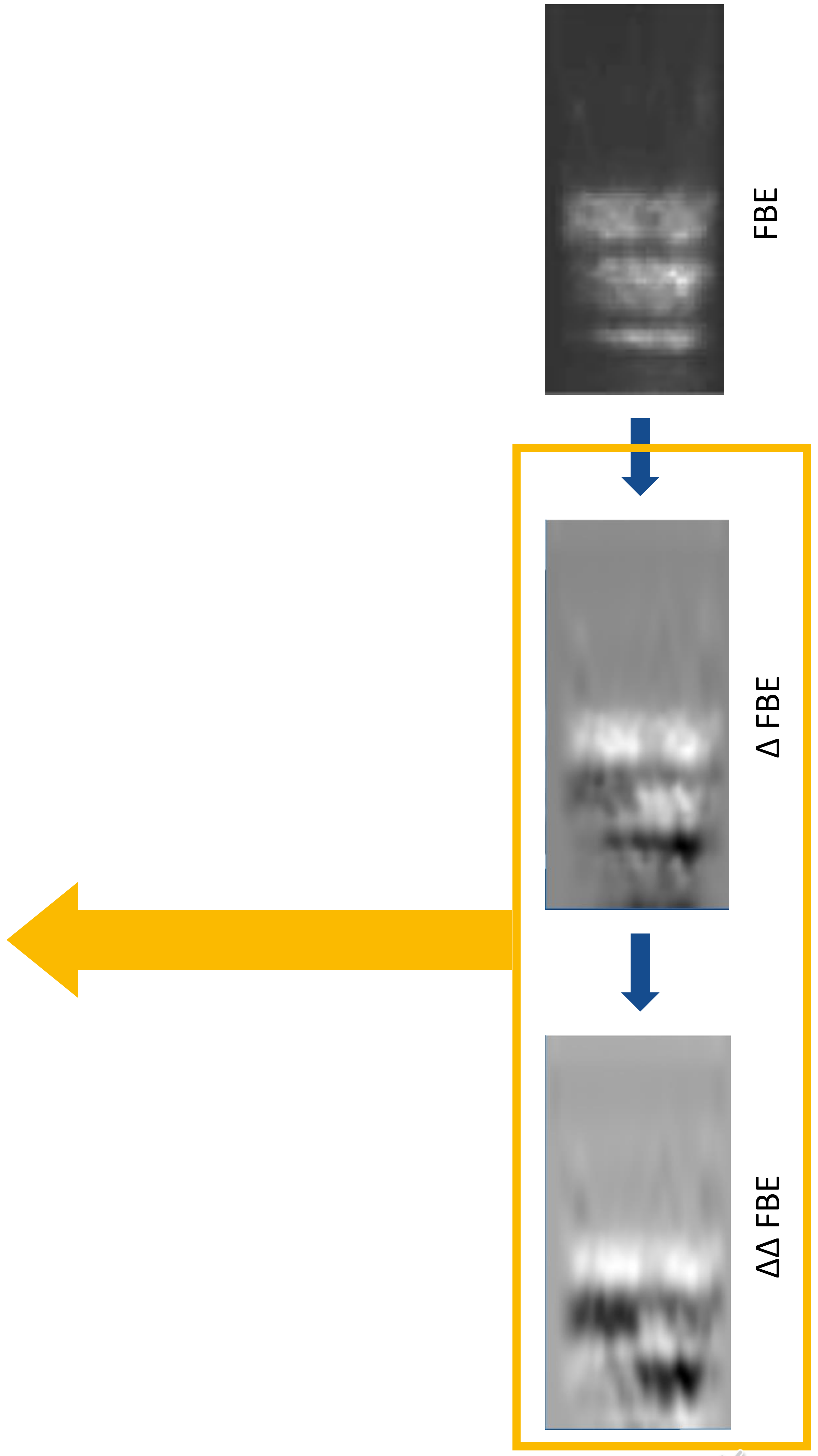
W = 3

Methodology

np.: FBE 64 x 30
+ ΔFBE 64 x 30

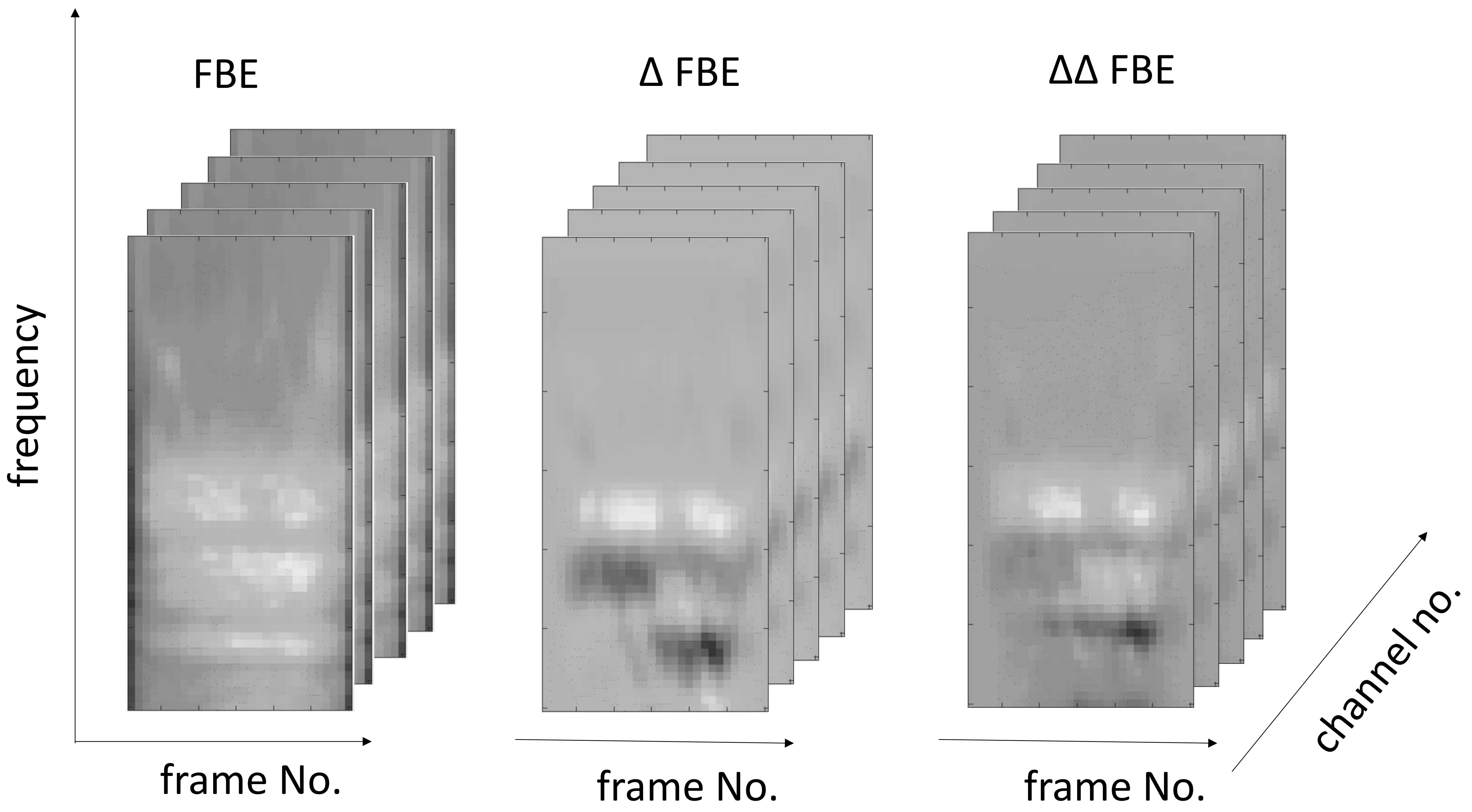


np.: FBE 64 x 30
+ ΔFBE 64 x 30
+ ΔΔFBE 64 x 30



W = 3

Methodology



INTERPOLATION OF FEATURE VALUES*

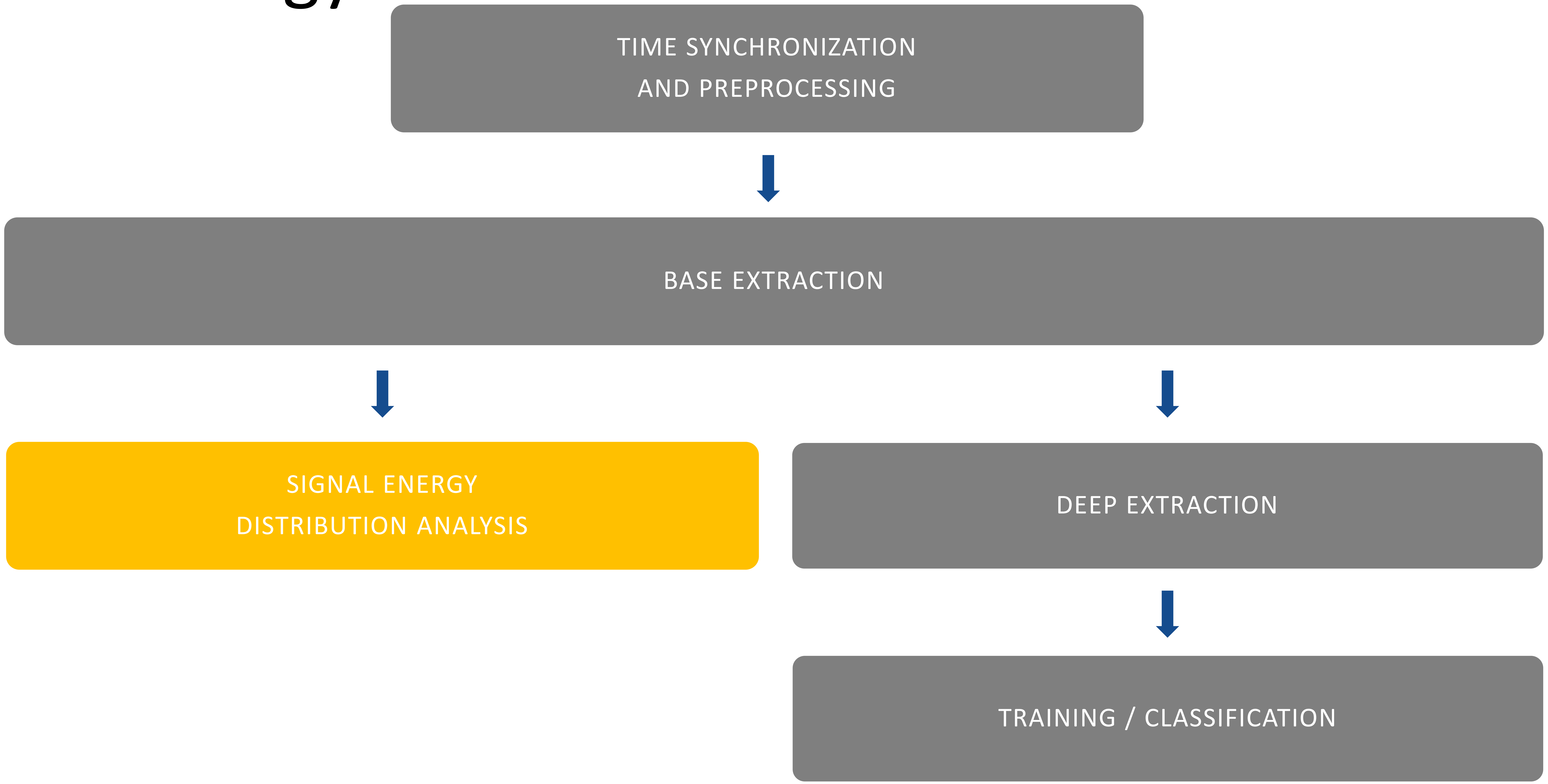
- NUMBER OF FRAMES = 64
- ONE-DIMENSIONAL LINEAR INTERPOLATION

FBE = CH5 x 64 x 64

DEEP EXTRACTION

* - independently for all features from a given channel

Methodology



Methodology

Class

normative realization for the sounds /s/ and /ʃ/

→ **klasa** s_{norm} **i** $\int_{norm}$

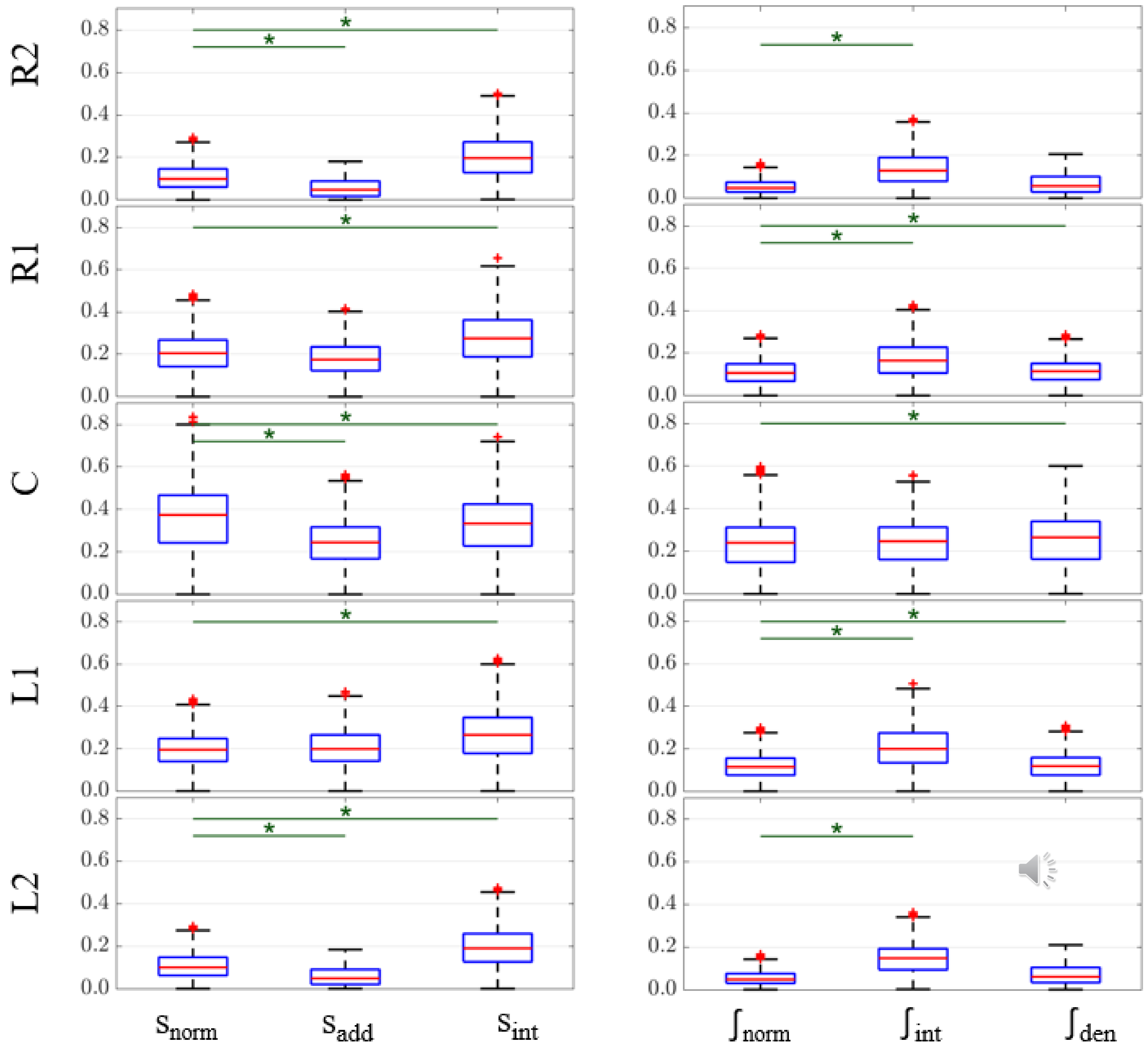
interdental and addental realization for the sound /s/

→ **klasa** s_{int} **i** s_{add}

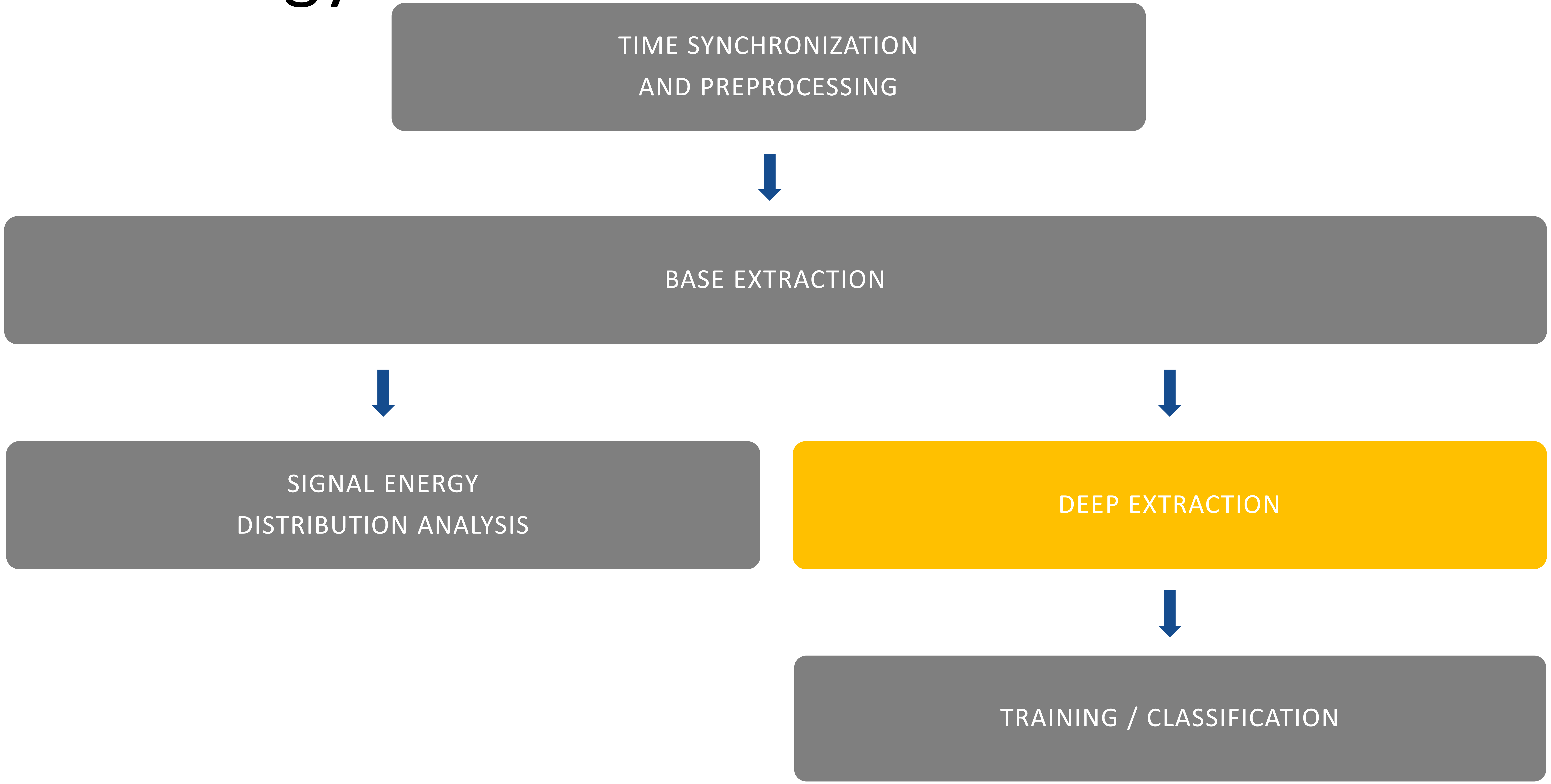
interdental and dental realization for the sound /ʃ/

→ **klasa** $\int_{int}$ **i** $\int_{den}$

* - significant difference at the level of $p < 0.05$

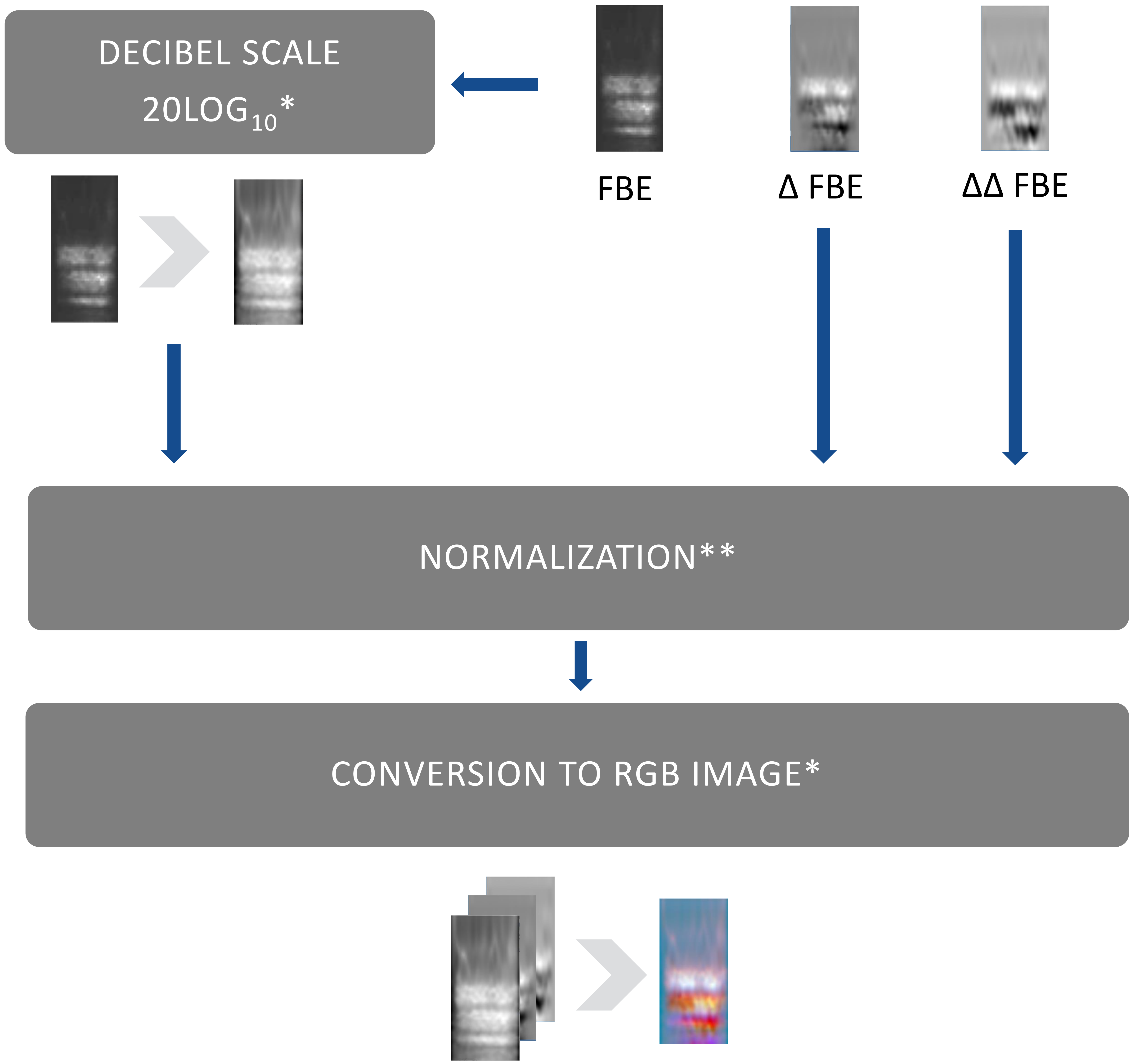


Methodology



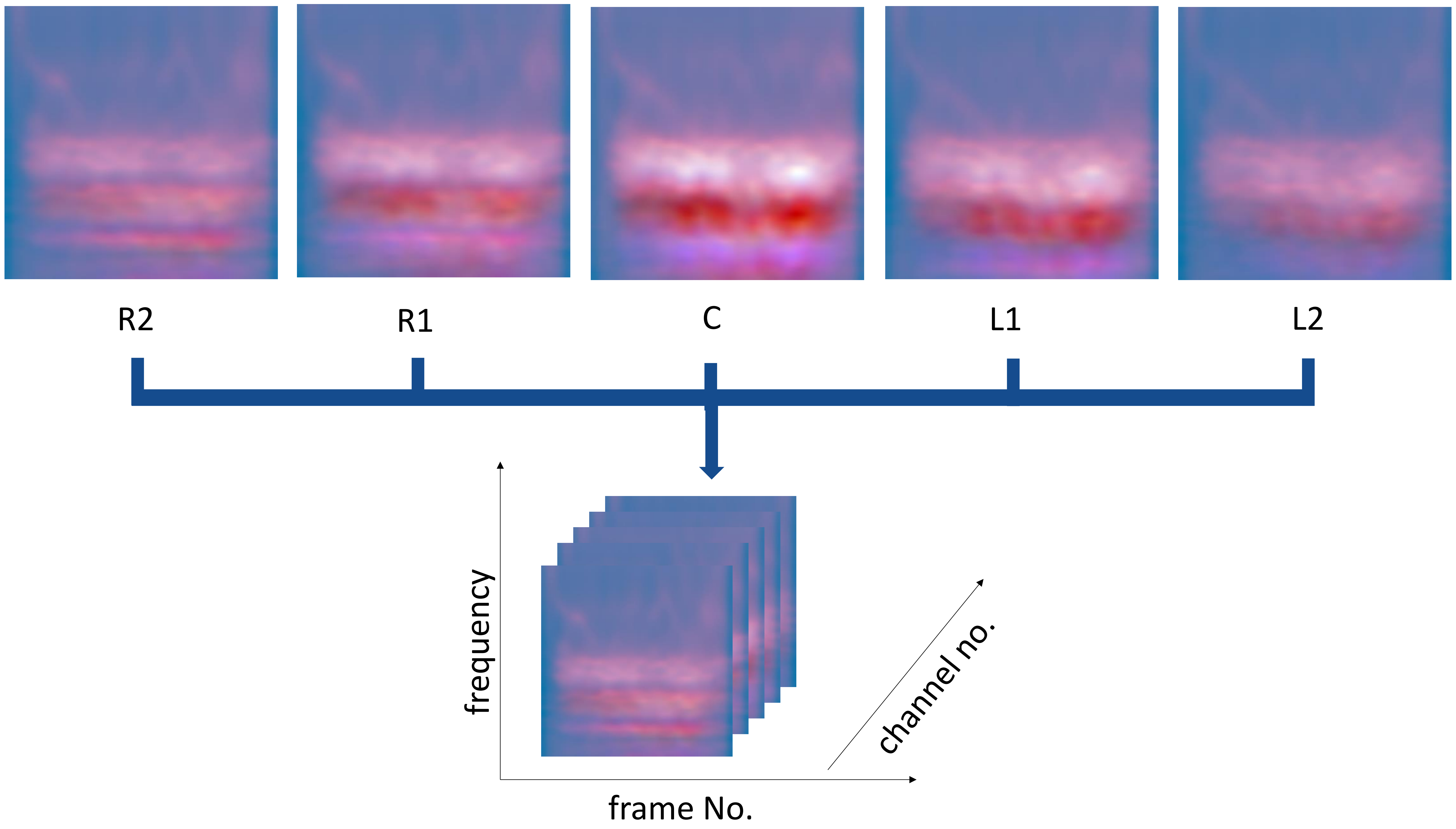
DEEP EXTRACTION - STEPS TO CONVERT FEATURES TO AN IMAGE

Methodology



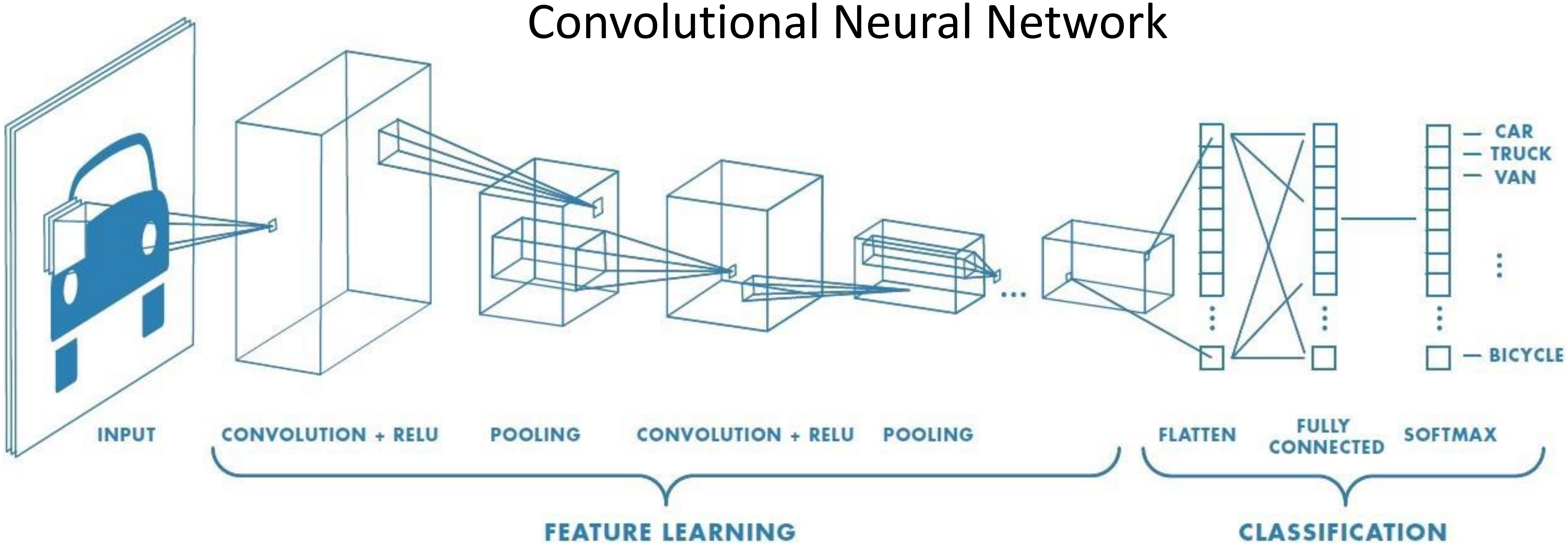
* - independently for each channel
** - based on all channels

Methodology



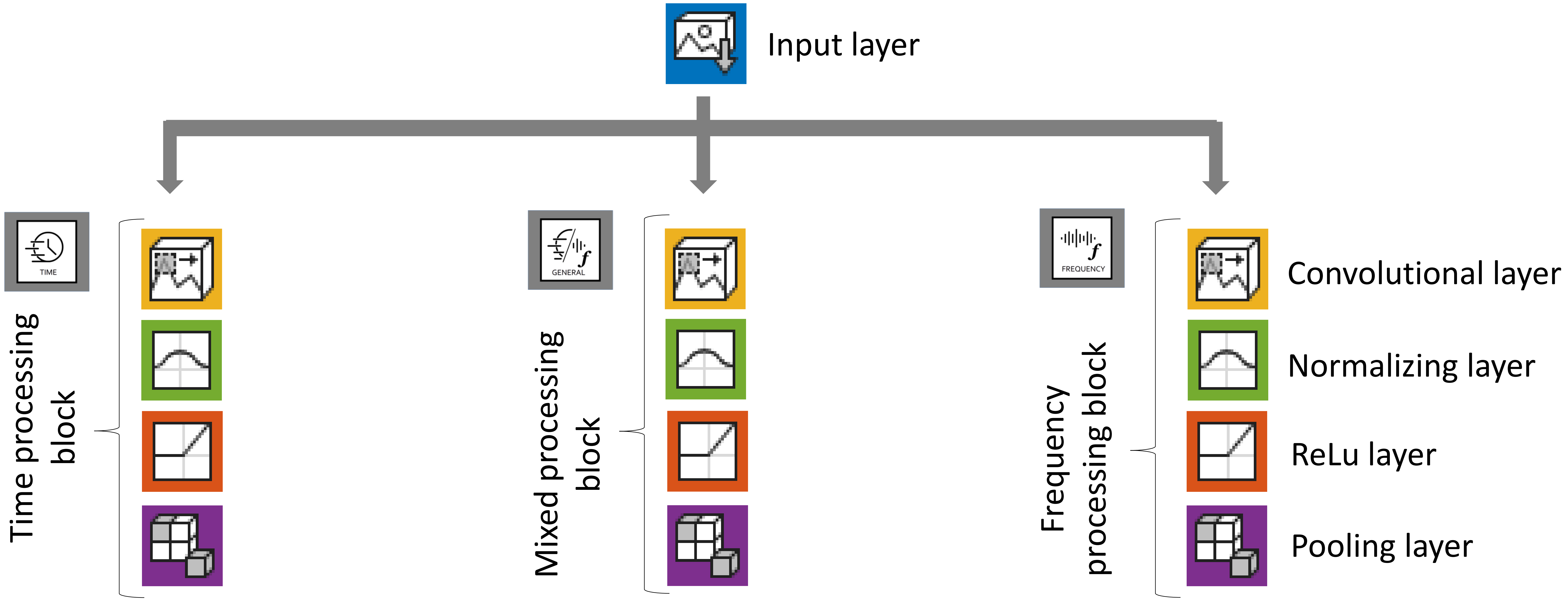
Methodology

Convolutional Neural Network

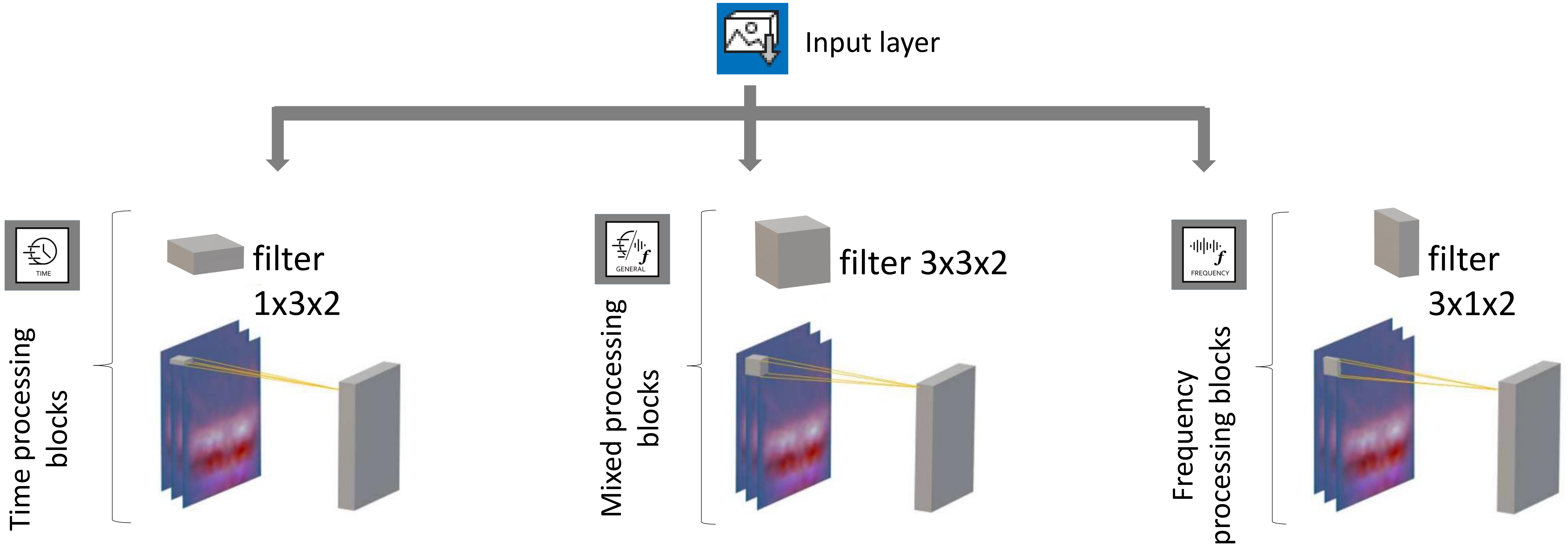


https://www.mathworks.com/

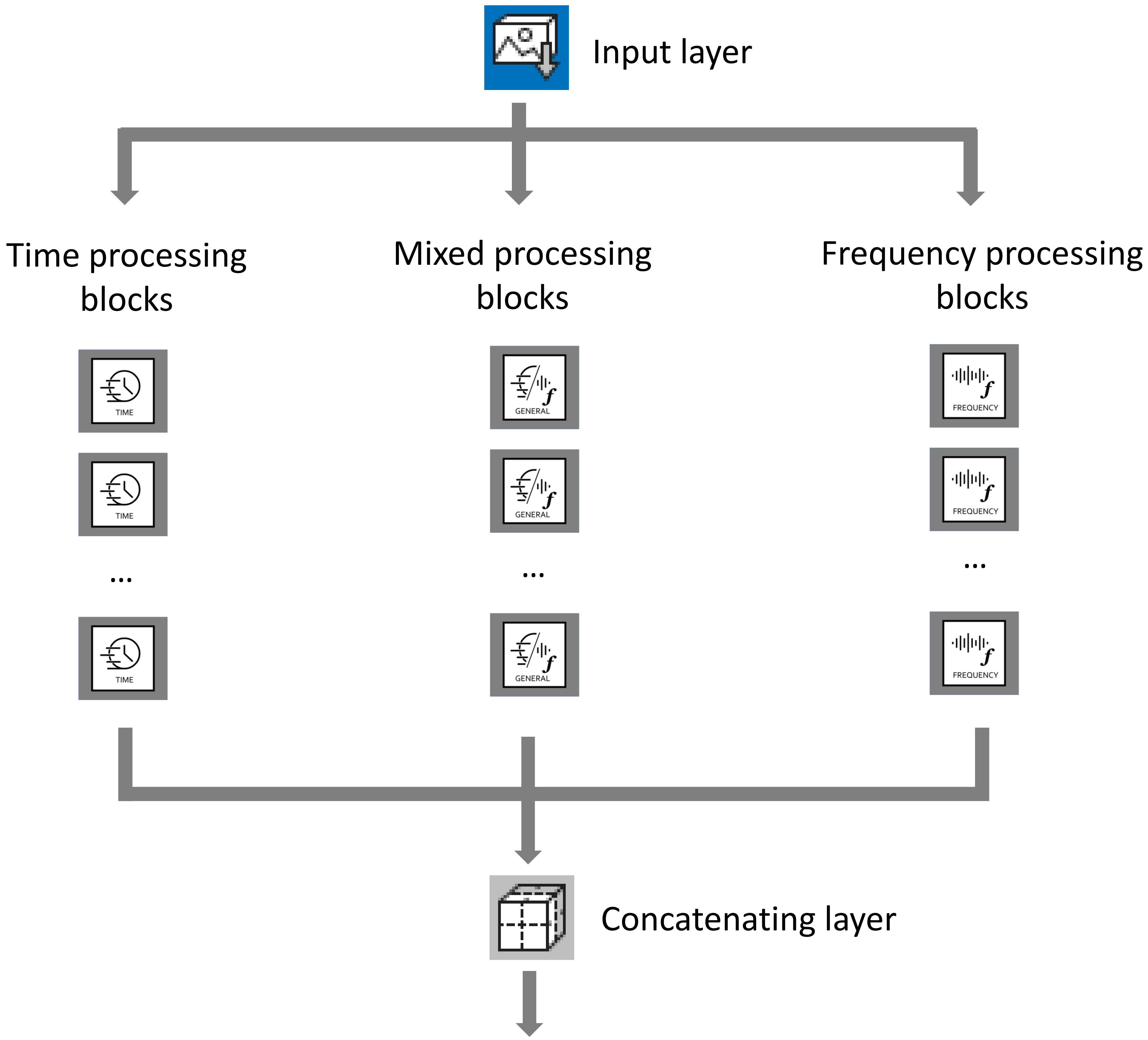
Methodology



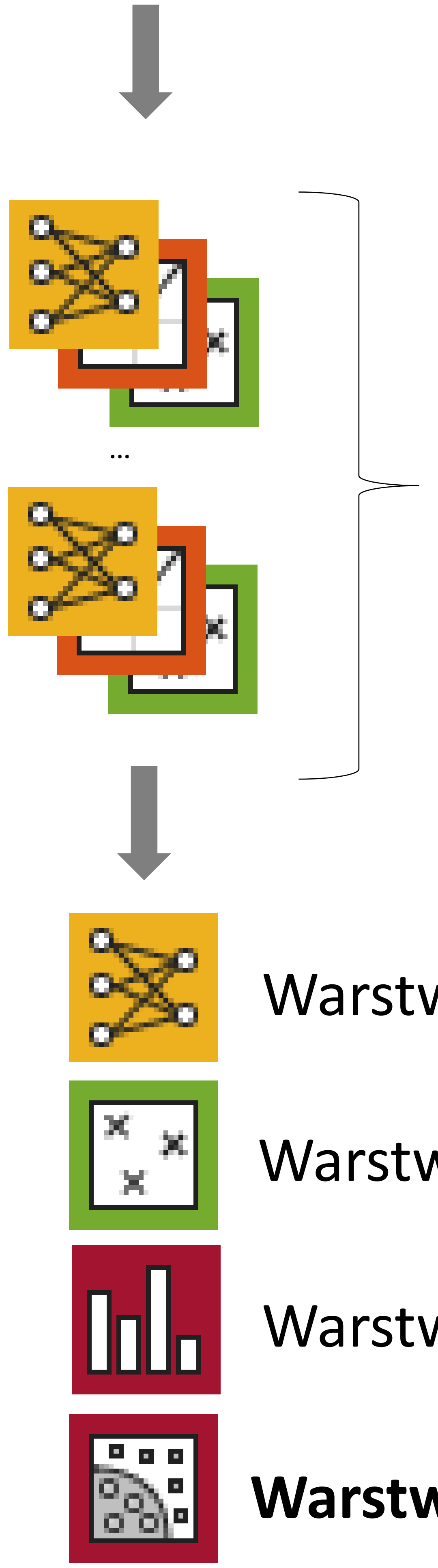
Methodology



Methodology



Methodology



Warstwy w pełni połączone
wraz z warstwami ReLu oraz
z warstwami typu dropout

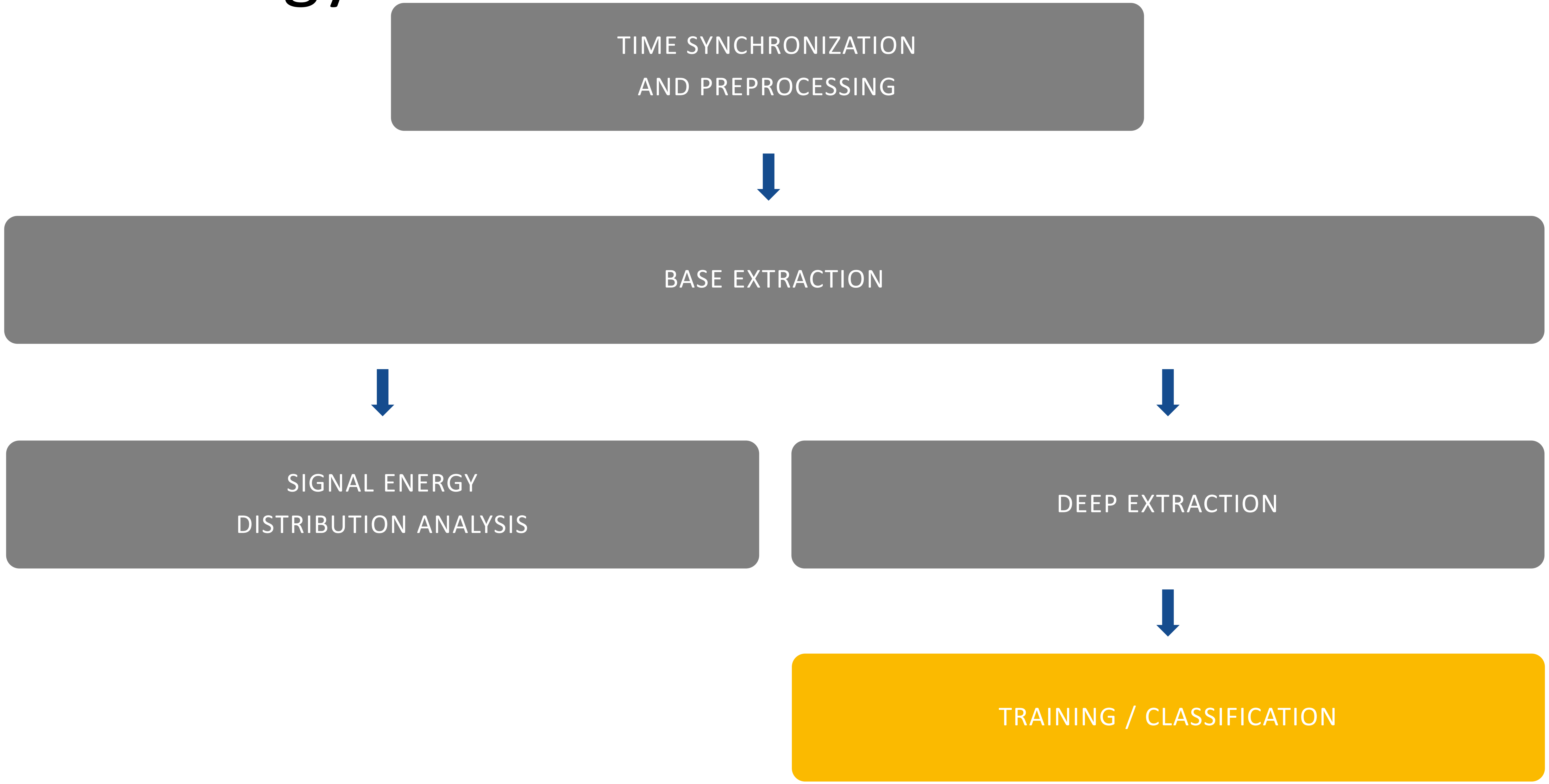
Warstwa w pełni połączona

Warstwa typu dropout

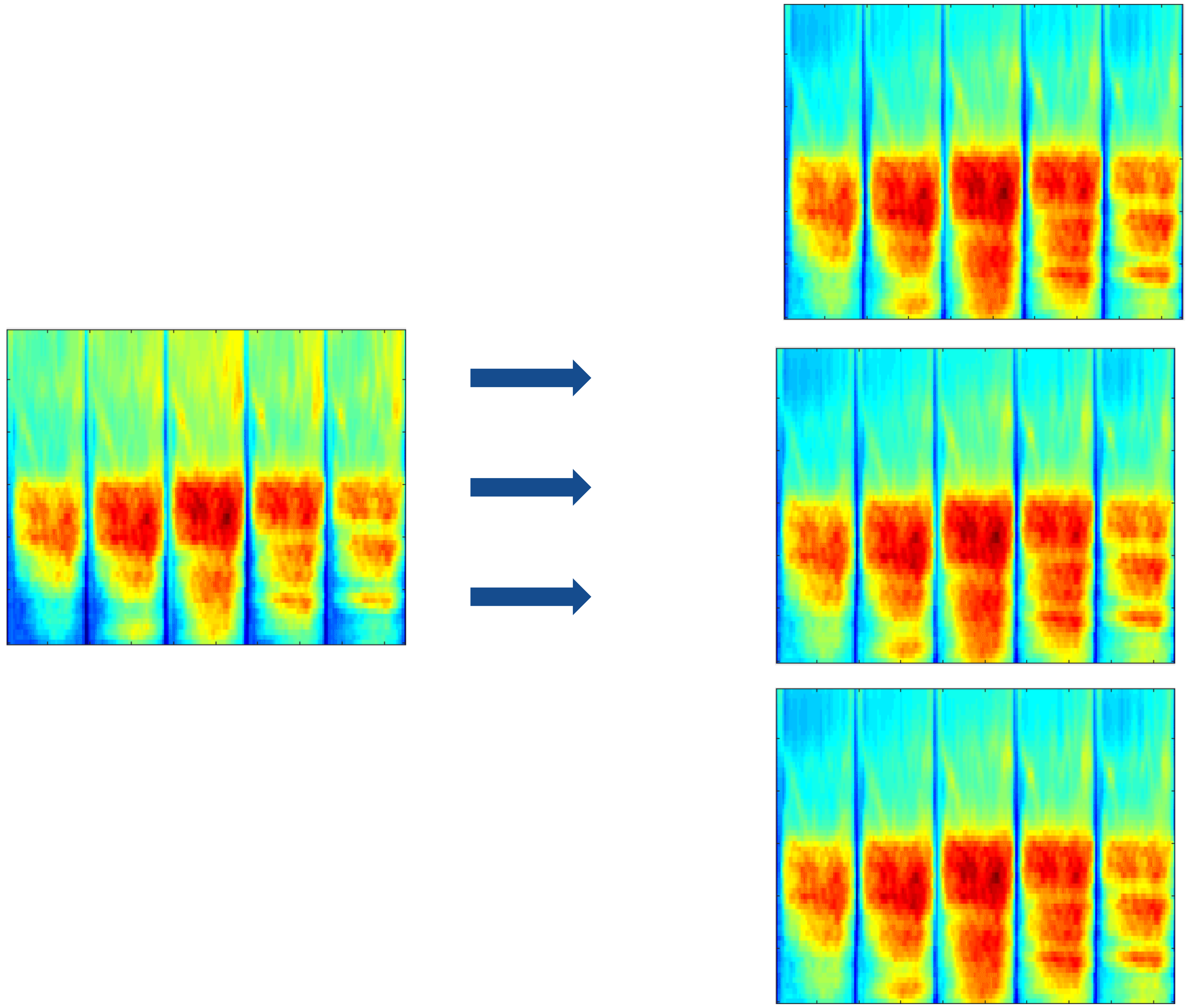
Warstwa typu softmax

Warstwa klasyfikująca

Methodology



Methodology

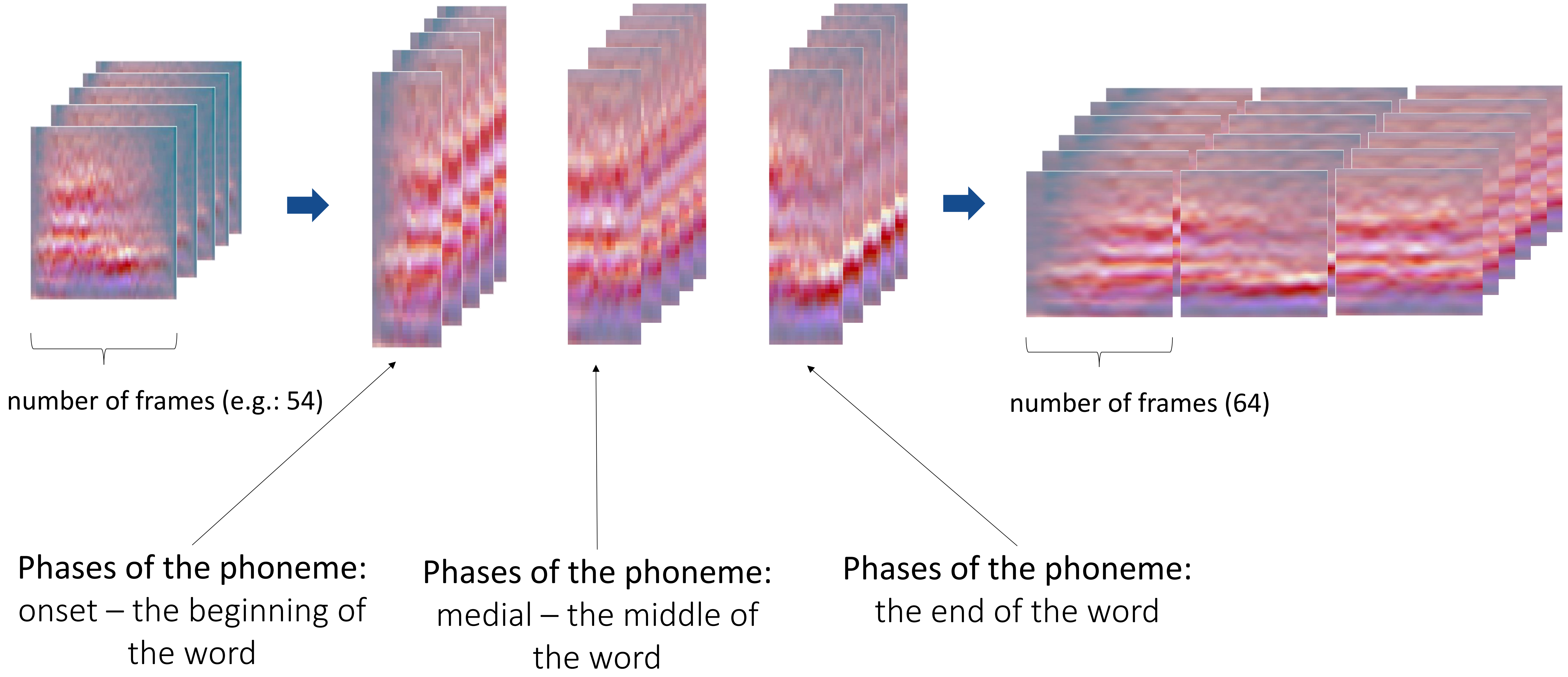


α = 0,90

α = 0,91

α = 0,92

Methodology



Danuta Ostaszewska, Jolanta Tambor: Fonetyka i fonologia współczesnego języka polskiego. Warszawa, 2000

Methodology

Network parameter name	Parameter value
filter size	3 x 1 x 2 1 x 3 x 2 3 x 3 x 2
input data size	64 x 64 x 5 (x 3)
model depth - number of processing blocks	5
model width - number of filters	16-64
training batch size	128

Summary of the parameters of the proposed model used during the experiments

Methodology

Network parameter name	Parameter value
filter size	3 x 1 x 2 1 x 3 x 2 3 x 3 x 2
input data size	64 x 64 x 5 (x 3)
model depth - number of processing blocks	5
model width - number of filters	16-64
training batch size	128

Learning process configuration	Hyperparameter value
optimization algorithms	SGDM
number of epochs	25
number of epochs to stop the learning process early	5
learning rate coefficient	0,001
learning rate update coefficient	0,1
number of epochs to change the learning rate	10
L2	0,01
momentum	0,95

Summary of parameters and hyperparameters of the proposed model used during the experiments

Experiments

ANALYSIS OF SIBILANTS MODELS

PHONEME CLASSIFICATION

SENSITIVITY ANALYSIS OF THE CNN MODEL

Experiments

ANALYSIS OF
SIBILANTS MODELS

PHONEME
CLASSIFICATION

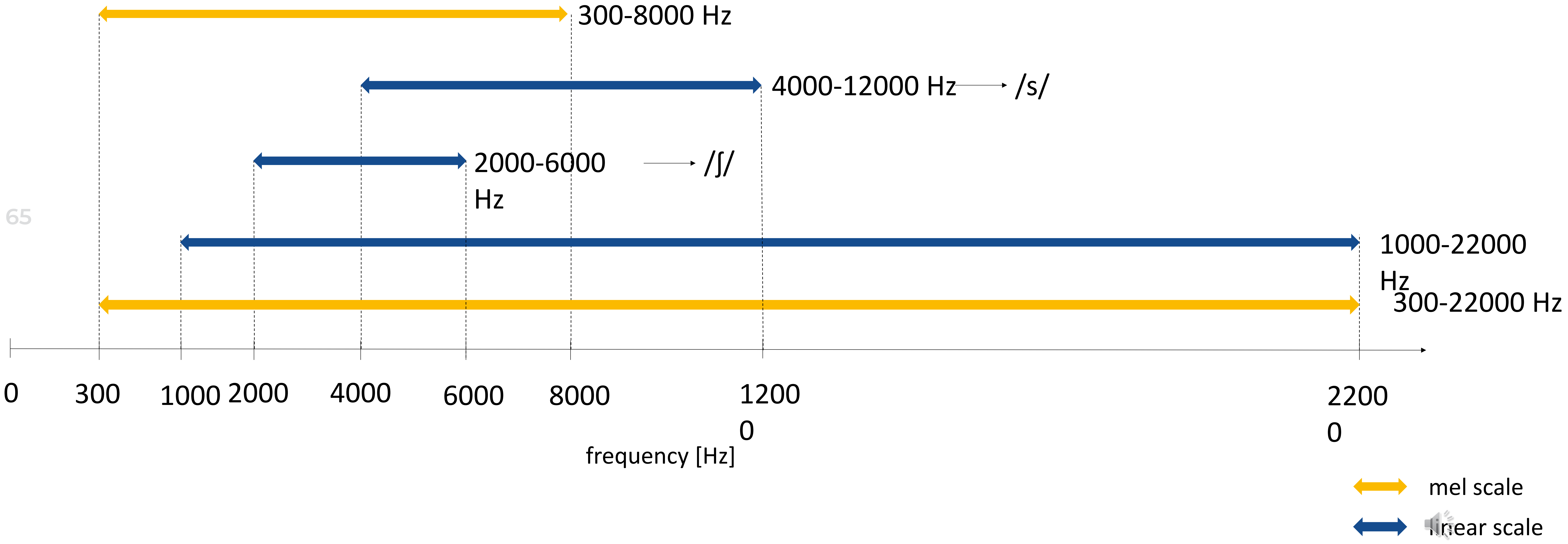
SENSITIVITY ANALYSIS OF THE CNN MODEL

Experiments

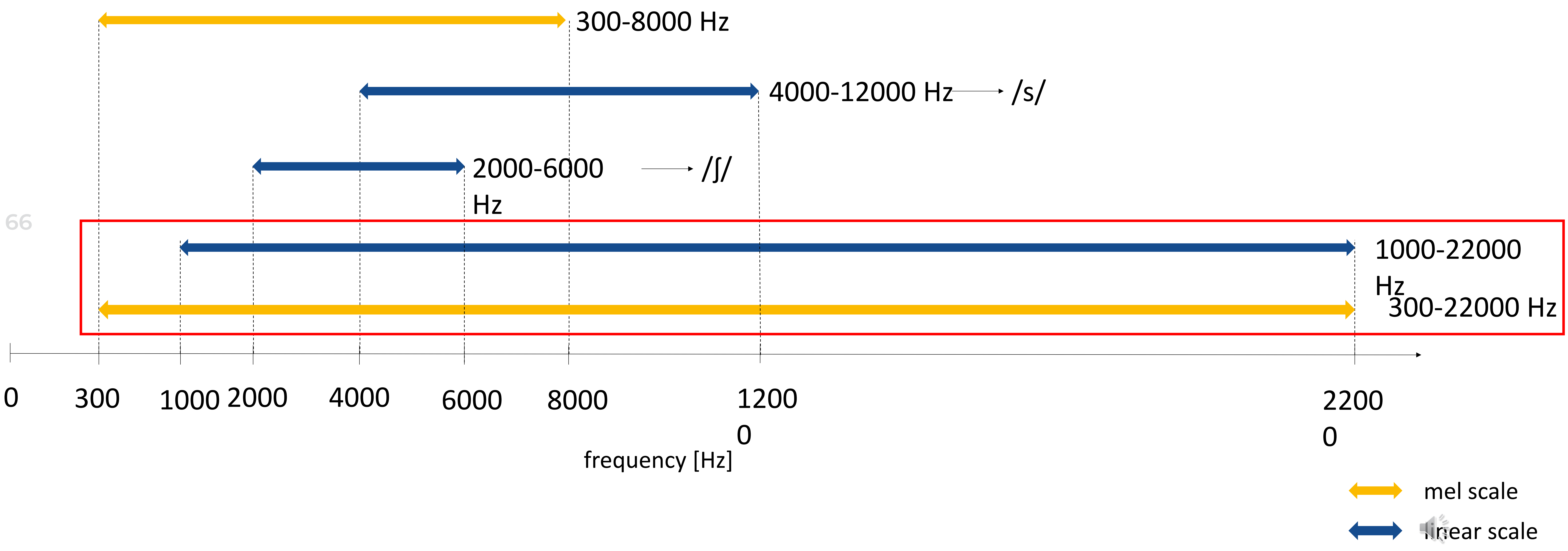
PHONEME CLASSIFICATION

- frequency range of the filter set,
- number of channels (1-CH vs 5-CH vs 15-CH),
- different variants of processing blocks (PB1 vs PB2).

Experiments



Experiments



Experiments

	Class	Dataset
normative realization for the sounds /s/ and /ʃ/	**klasa** s_{norm} **i** $ʃ_{norm}$	$\{ s, ʃ \}_2$
normative, interdental and addental realization for the sound /s/	**klasa** s_{norm} , s_{int} **i** s_{add}	$\{ s \}_3$
normative, interdental and dental realization for the sound /ʃ/	**klasa** $ʃ_{norm}$, $ʃ_{int}$ **i** $ʃ_{den}$	$\{ ʃ \}_3$
normative and non-normative realization for the sounds /s/ and /ʃ/	**klasa** s_{norm} , s_{int} **i** s_{add}	$\{ s, ʃ \}_6$
	$ʃ_{norm}$, $ʃ_{int}$ **i** $ʃ_{den}$	$\{ s, ʃ \}_4$

PHONEME CLASSIFICATION - CLASSES AND DATA SETS

Experiments

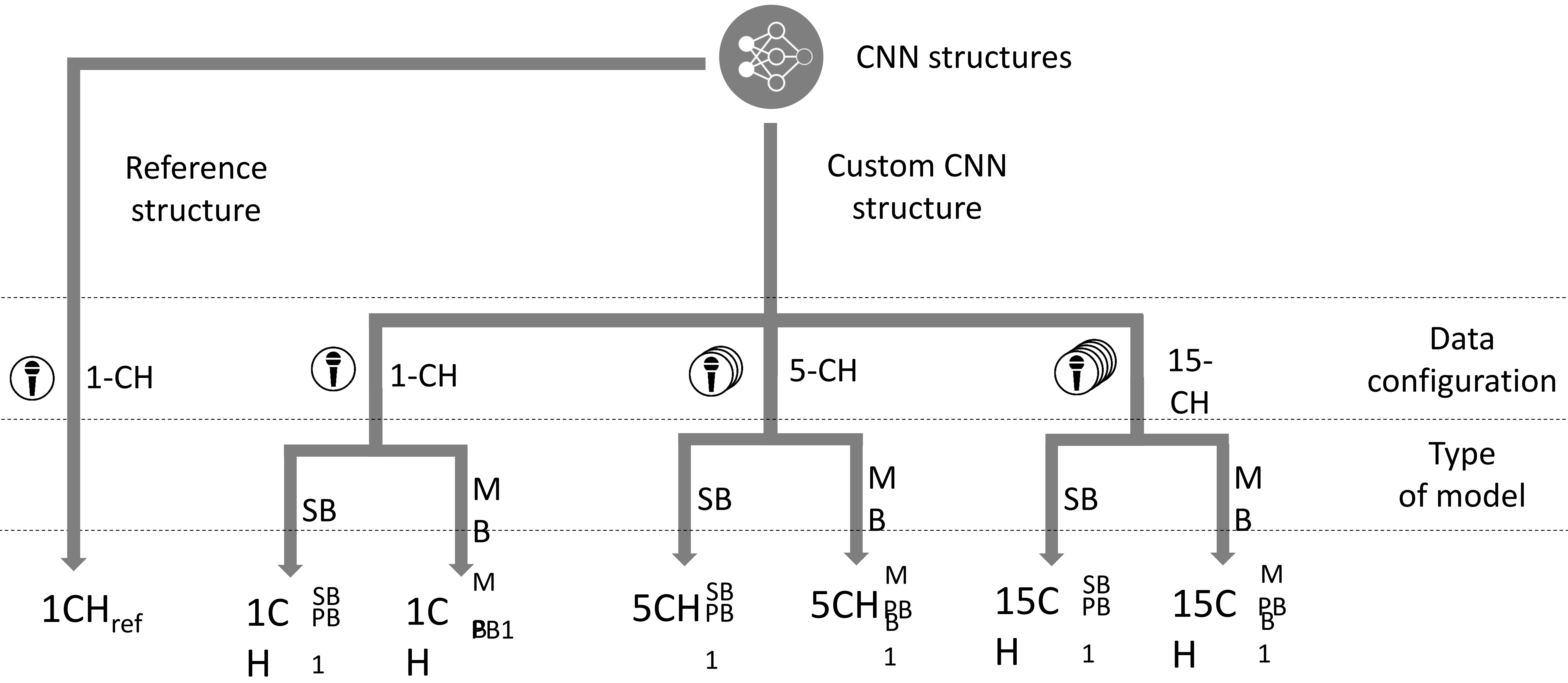
	Class	Dataset
normative realization for the sounds /s/ and /ʃ/	**klasa** s_{norm} **i** ʃ_{norm}	$\{\, s, \text{ʃ} \,\}_2$
normative, interdental and addental realization for the sound /s/	**klasa** s_{norm} **,** s_{int} **i** s_{add}	$\{\, s \,\}_3$
normative, interdental and dental realization for the sound /ʃ/	**klasa** ʃ_{norm} **,** ʃ_{int} **i** ʃ_{den}	$\{\, \text{ʃ} \,\}_3$
normative and non-normative realization for the sounds /s/ and /ʃ/	**klasa** s_{norm} **,** s_{int} **i** s_{add}	$\{\, s, \text{ʃ} \,\}_6$
	ʃ_{norm} **,** ʃ_{int} **i** ʃ_{den}	$\{\, s, \text{ʃ} \,\}_4$

Experiments

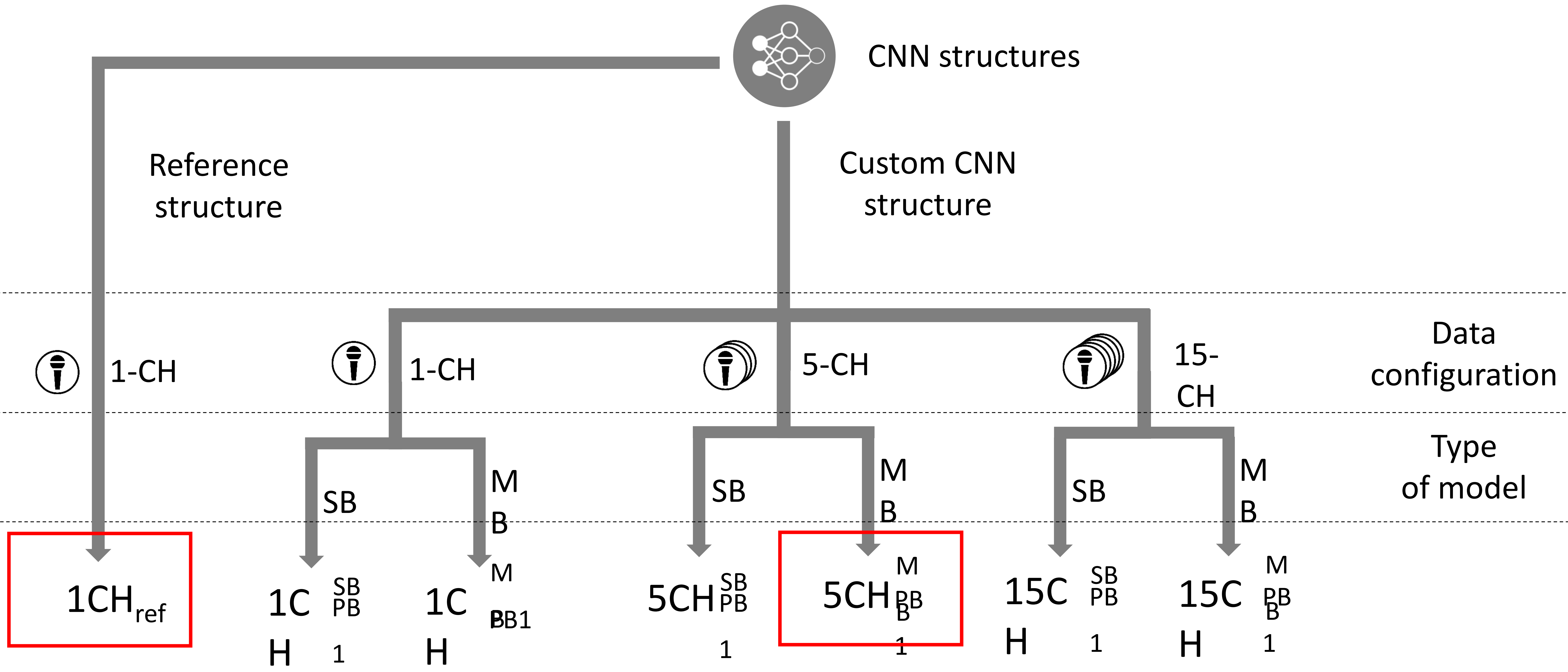
	Type of realisation	/s/			/ʃ/		
		normative	addentalna	interdentalna	normative	dentalna	interdentalna
Type of set	training 70%	151 (906)	64 (384)	116 (696)	163 (978)	54 (324)	84 (504)
	validation 15%	33	14	26	35	12	18
	test 15%	33	14	26	35	12	18

The size of the dataset (before and after augmentation) in recognizing different realizations of the sounds /s/ and /ʃ/

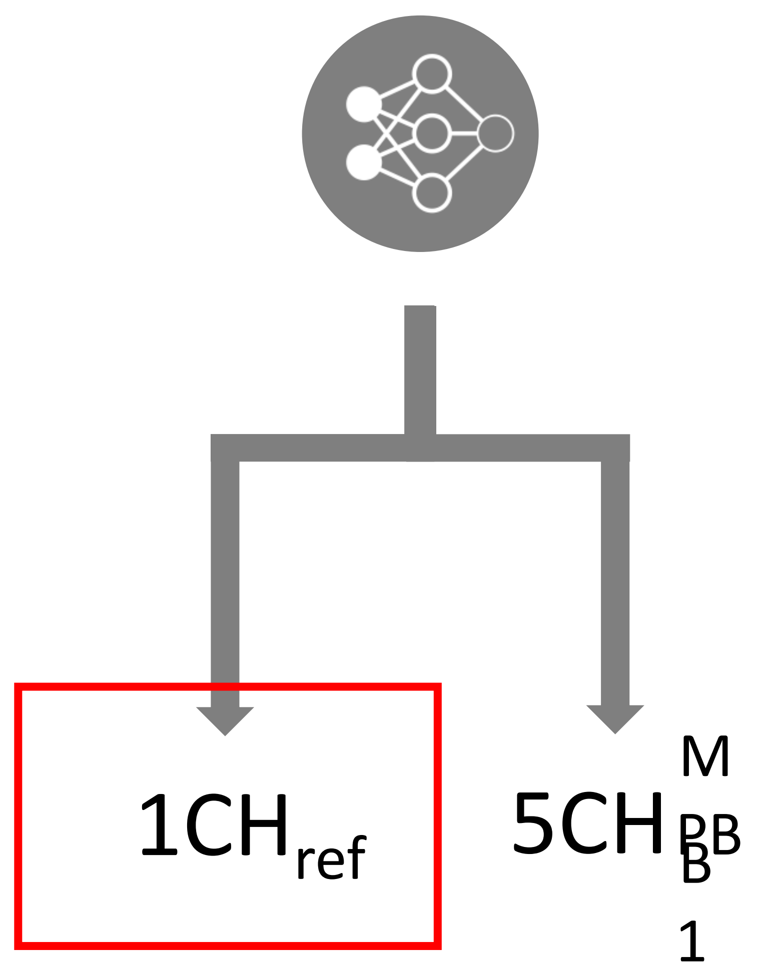
Experiments

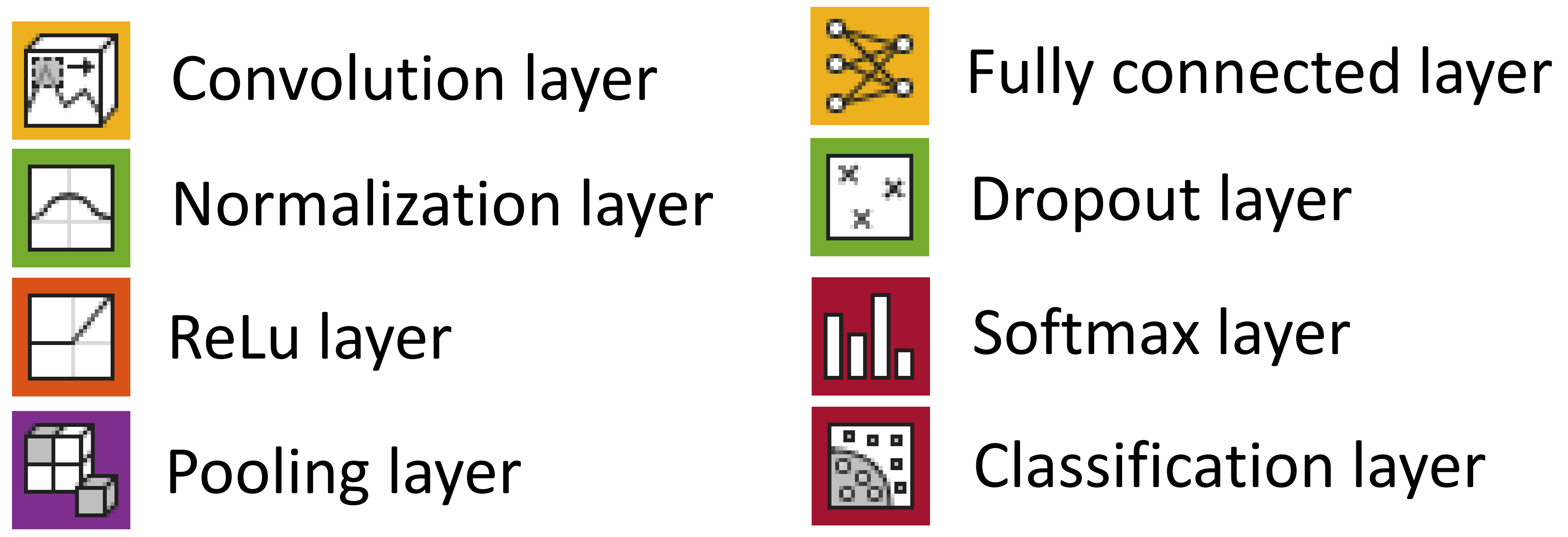


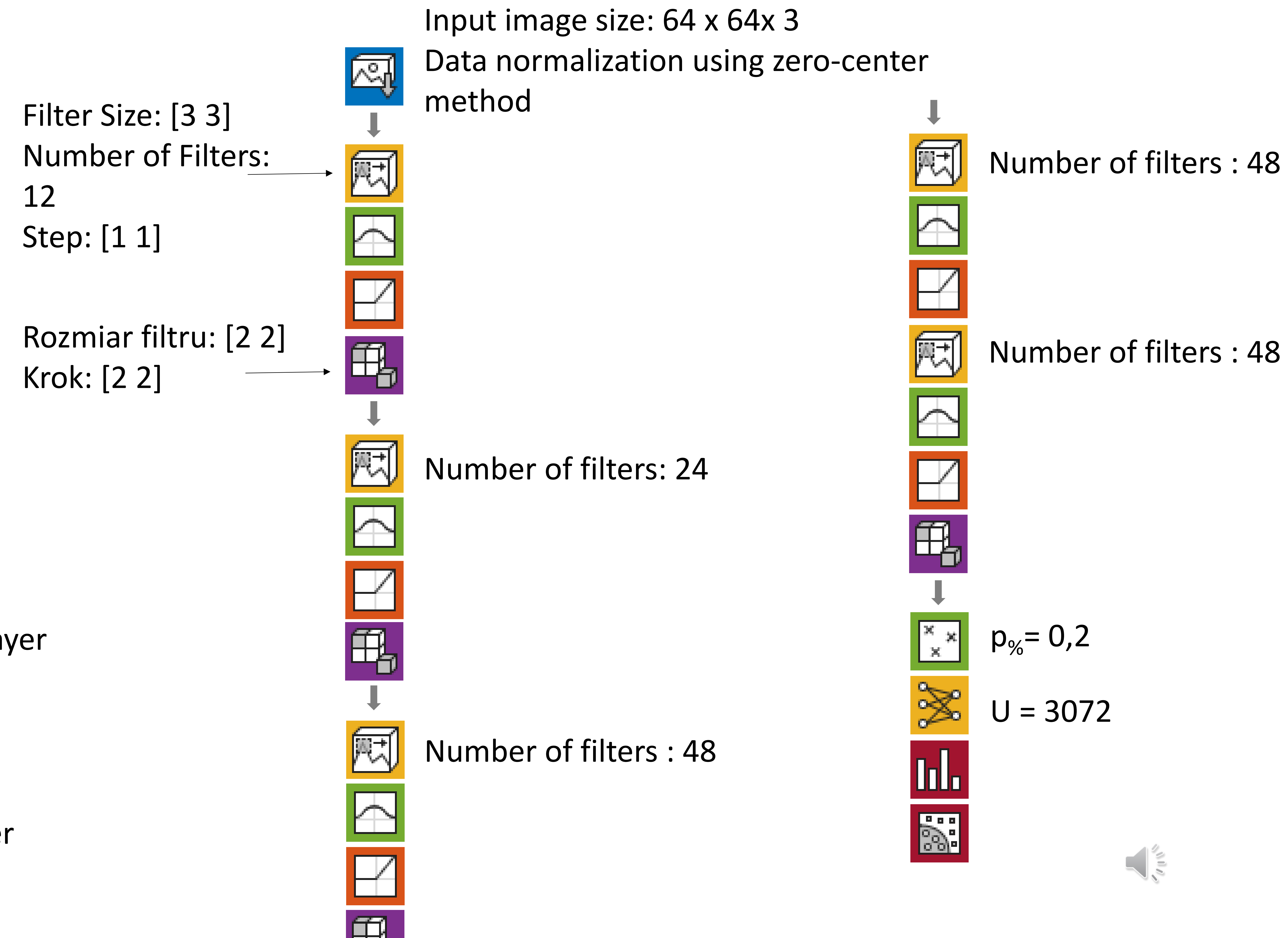
Experiments



Experiments

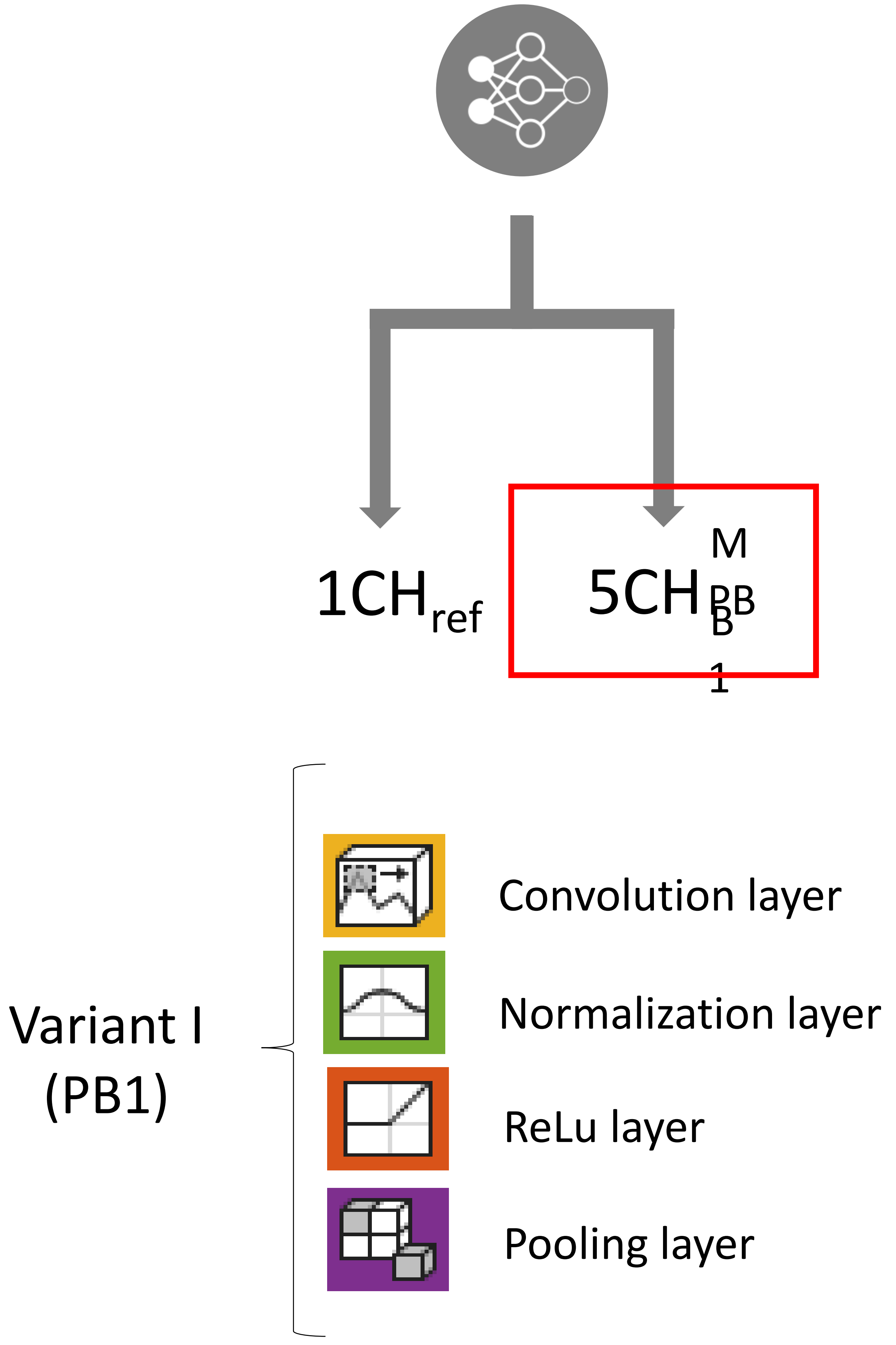


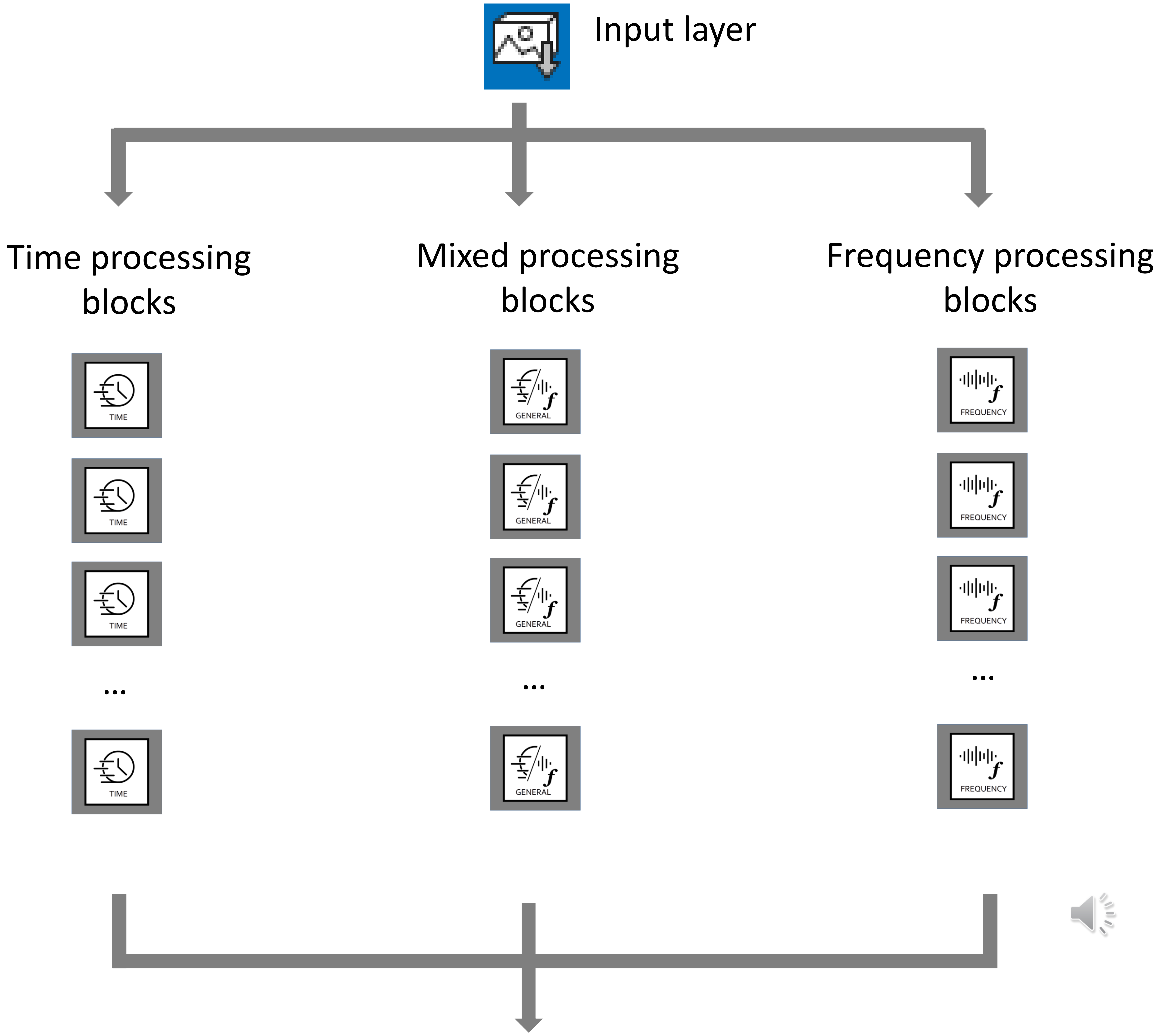




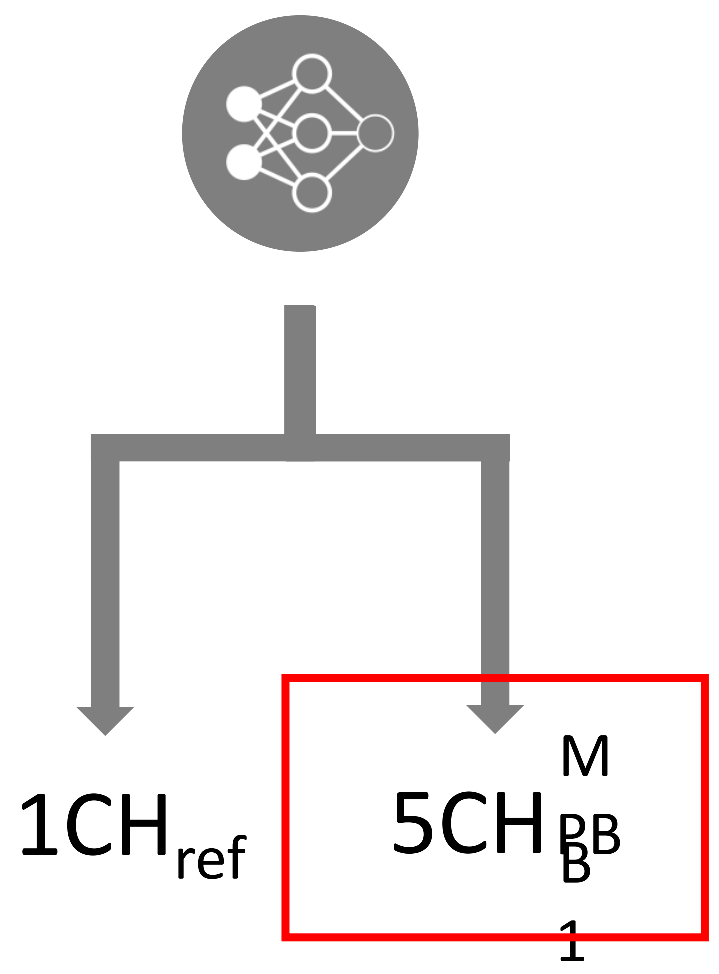
Warden P. "Speech Commands: A public dataset for single-word speech recognition", 2017

Experiments

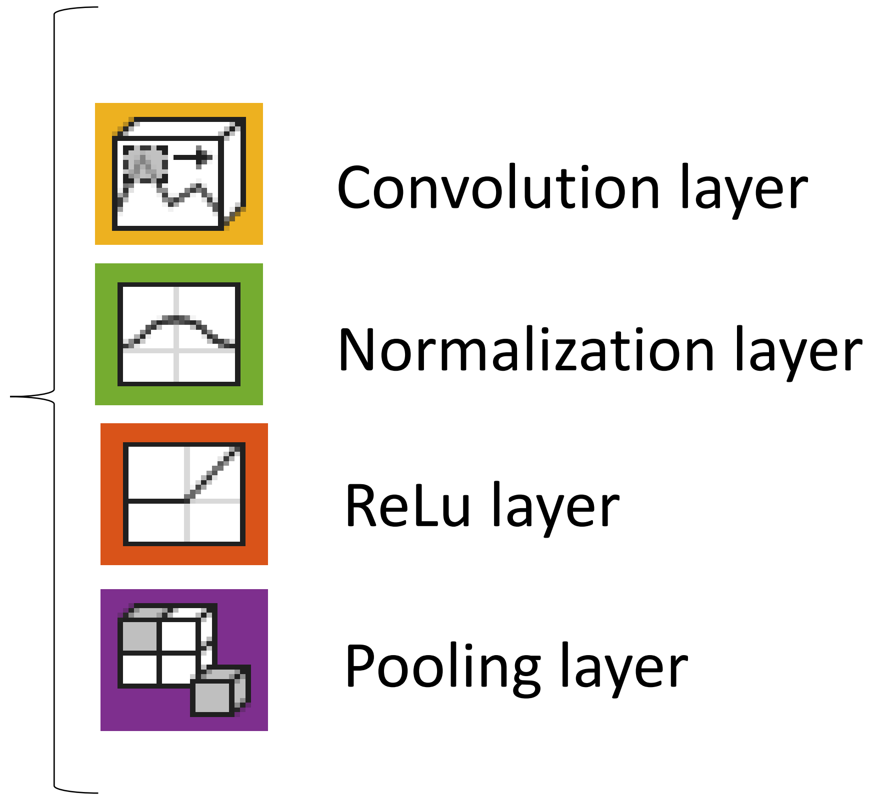


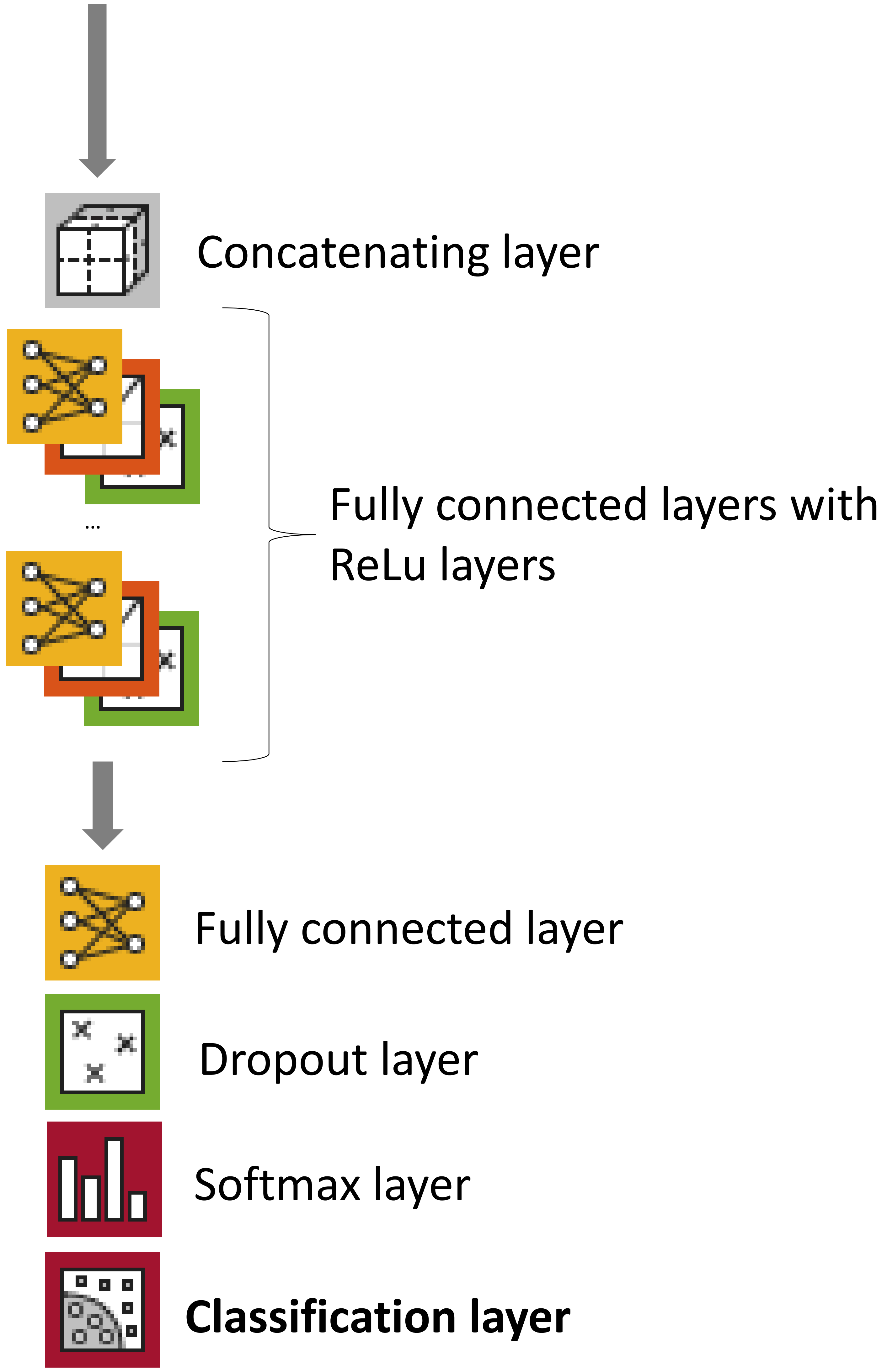


Experiments



Variant I
(PB1)





Experiments

$1CH_{ref}$

[1000-22000 Hz]

Reference \ Prediction	s_{norm}	s_{add}	s_{int}	$ʃ_{norm}$	$ʃ_{int}$	$ʃ_{den}$	PPV
s_{norm}	289	46	140	5	7	173	43.8%
s_{add}	92	23	55	8	5	97	8.2%
s_{int}	112	24	135	35	37	176	26.0%
$ʃ_{norm}$	1		19	516	67	92	74.2%
$ʃ_{int}$	9	6	33	115	22	52	9.3%
$ʃ_{den}$	61	5	54	63	20	157	43.6%
TPR	51.2%	22.1%	31.0%	69.5%	13.9%	21.0%	

ACC: 41.5%
TPR: 34.1%
TNR: 88.1%

$5CH_{MBBB1}$

[1000-22000 Hz]

Reference \ Prediction	s_{norm}	s_{add}	s_{int}	$ʃ_{norm}$	$ʃ_{int}$	$ʃ_{den}$	PPV
s_{norm}	414	47	59	19	1	120	62.7%
s_{add}	69	159	8	8	2	34	56.8%
s_{int}	36	5	374	13	57	34	72.1%
$ʃ_{norm}$	10	5	21	551	46	62	79.3%
$ʃ_{int}$		1	95	50	78	13	32.9%
$ʃ_{den}$	88	12	43	92	24	101	28.1%
TPR	67.1%	69.4%	62.3%	75.2%	37.5%	27.7%	

ACC: 60.9%
TPR: 55.3%
TNR: 92.0%

Experiments

1CH$_{ref}$

[1000-22000 Hz]

Reference \ Prediction	s_{norm}	s_{add}	s_{int}	ʃ$_{norm}$	ʃ$_{int}$	ʃ$_{den}$	PPV
s_{norm}	289	46	140	5	7	173	43.8%
s_{add}	92	23	55	8	5	97	8.2%
s_{int}	112	24	135	35	37	176	26.0%
ʃ$_{norm}$	1		19	516	67	92	74.2%
ʃ$_{int}$	9	6	33	115	22	52	9.3%
ʃ$_{den}$	61	5	54	63	20	157	43.6%
TPR	51.2%	22.1%	31.0%	69.5%	13.9%	21.0%	

ACC: 41.5%
TPR: 34.1%
TNR: 88.1%

5CH$^{M}_{BB}$ B 1

[1000-22000 Hz]

Reference \ Prediction	s_{norm}	s_{add}	s_{int}	ʃ$_{norm}$	ʃ$_{int}$	ʃ$_{den}$	PPV
s_{norm}	414	47	59	19	1	120	62.7%
s_{add}	69	159	8	8	2	34	56.8%
s_{int}	36	5	374	13	57	34	72.1%
ʃ$_{norm}$	10	5	21	551	46	62	79.3%
ʃ$_{int}$		1	95	50	78	13	32.9%
ʃ$_{den}$	88	12	43	92	24	101	28.1%
TPR	67.1%	69.4%	62.3%	75.2%	37.5%	27.7%	

ACC: 60.9%
TPR: 55.3%
TNR: 92.0%

Experiments

$5CH^{M}_{BB}$ 1

[1000-22000 Hz]

Reference \ Prediction	s_{norm}	s_{add}	s_{int}	$ʃ_{norm}$	$ʃ_{int}$	$ʃ_{den}$	PPV
s_{norm}	414	47	59	19	1	120	62.7%
s_{add}	69	159	8	8	2	34	56.8%
s_{int}	36	5	374	13	57	34	72.1%
$ʃ_{norm}$	10	5	21	551	46	62	79.3%
$ʃ_{int}$		1	95	50	78	13	32.9%
$ʃ_{den}$	88	12	43	92	24	101	28.1%
TPR	67.1%	69.4%	62.3%	75.2%	37.5%	27.7%	

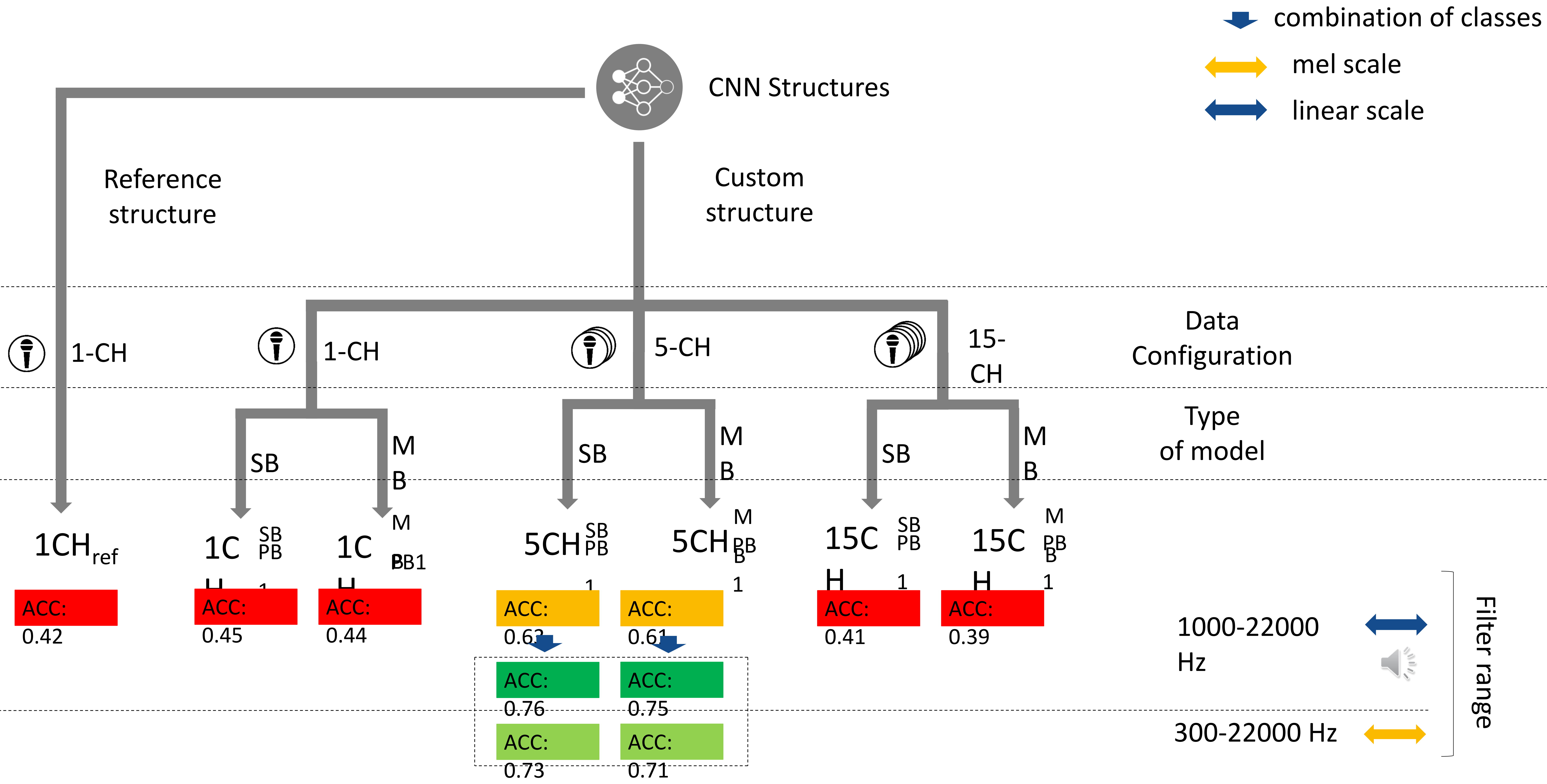
ACC: 60.9%
TPR: 55.3%
TNR: 92.0%

[1000-22000 Hz]

Reference \ Prediction	s_{norm} / $ʃ_{den}$	s_{add}	s_{int} / $ʃ_{int}$	$ʃ_{norm}$	PPV
s_{norm} / $ʃ_{den}$	757	78	125	59	74.3%
s_{add}	79	178	13	9	63.8%
s_{int} / $ʃ_{int}$	85	11	636	41	82.3%
$ʃ_{norm}$	63	7	93	529	76.4%
TPR	76.9%	65.0%	73.4%	82.9%	

ACC: 76.0%
TPR: 74.2%
TNR: 91.5%

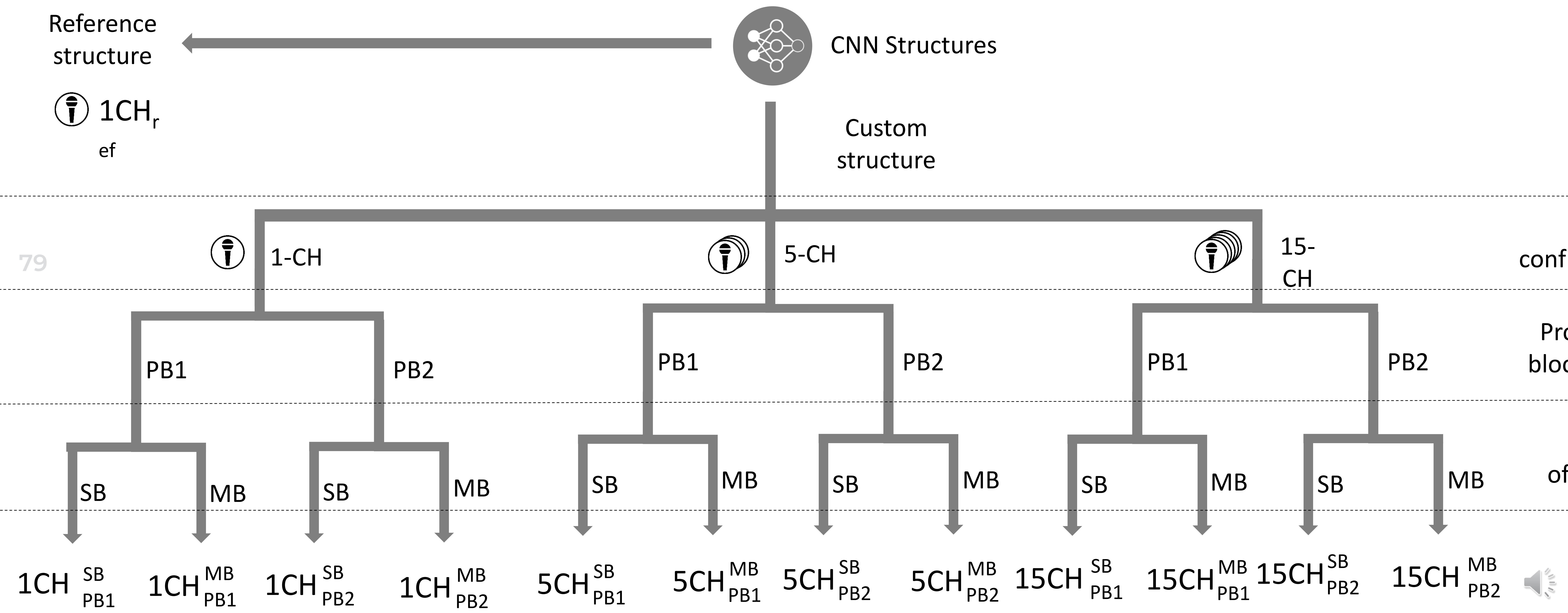
Experiments



Experiments

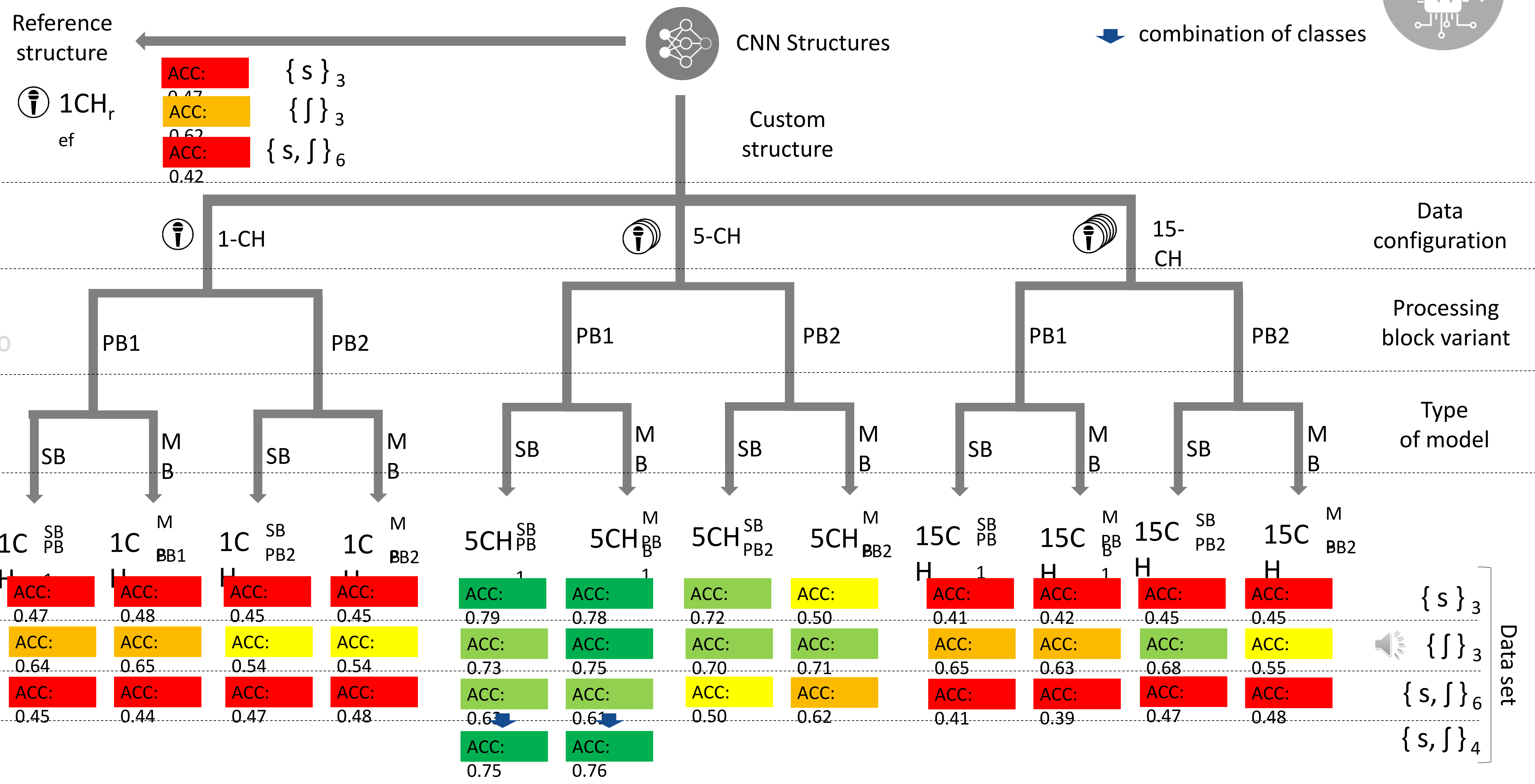
linear scale filter range: 1000-22000 Hz

combination of classes



PHONEM CLASSIFICATION – RESULTS OF TESTED CNN STRUCTURES

linear scale filter range: 1000-22000 Hz



PHONEM CLASSIFICATION – RESULTS OF TESTED CNN STRUCTURES

81

Experiments

ANALYSIS OF SIBILANTS MODELS

PHONEME CLASSIFICATION

SENSITIVITY ANALYSIS OF THE CNN MODEL

Experiments

ANALYSIS OF SIBILANTS MODELS

PHONEME CLASSIFICATION

SENSITIVITY ANALYSIS OF THE CNN MODEL

Experiments

SENSITIVITY ANALYSIS OF THE CNN MODEL

- the influence of selected hyperparameters and network structure on classification results,
- verification of the validity of network structure parallelization.

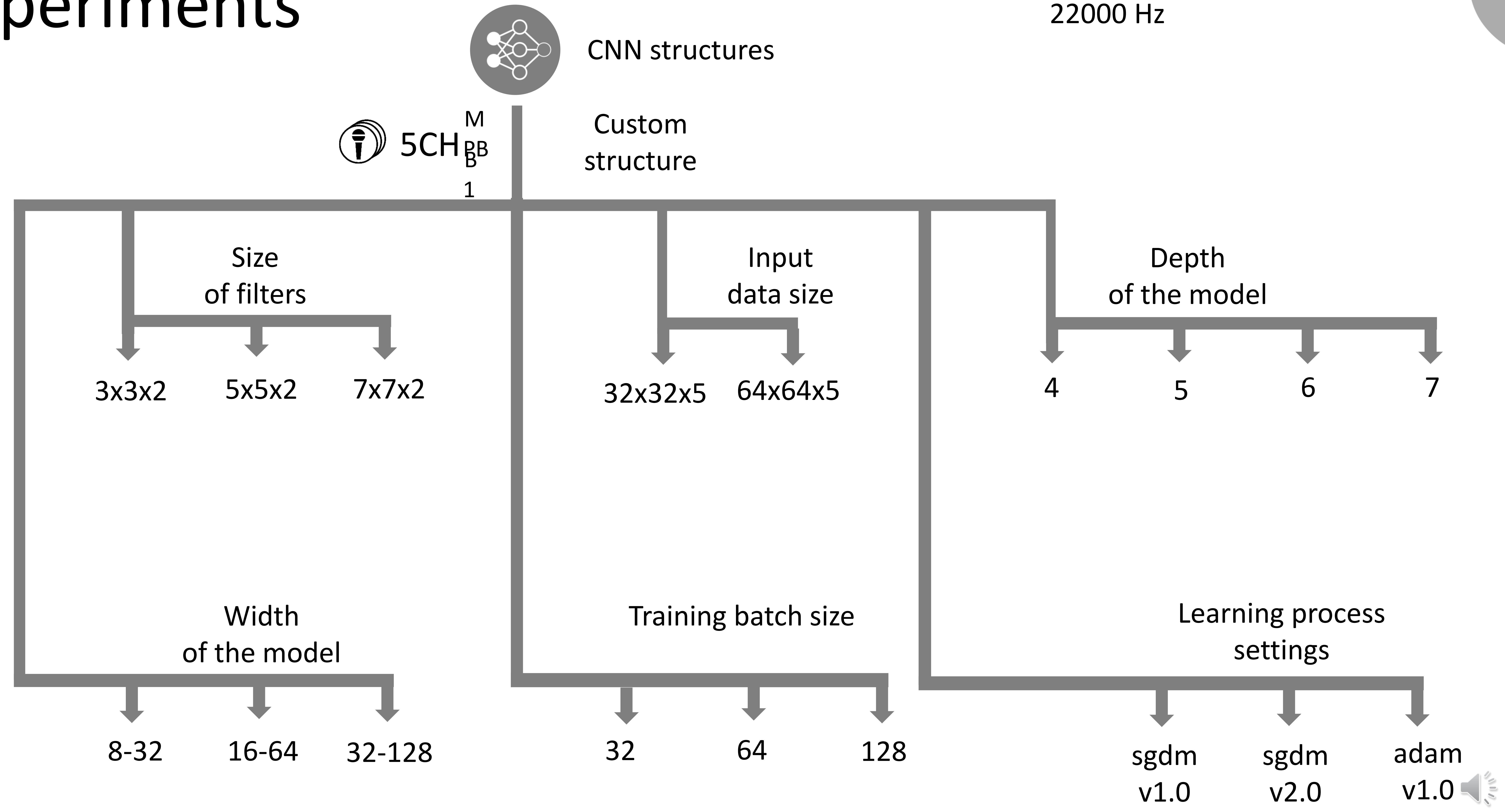
Experiments

	Type of realisation	/s/	/ʃ/	/s/ + /ʃ/	/s/
		normative	normative	interdentalna	addentalna
Type of set	training 70%	235 (1410)	163 (978)	118 (708)	116 (696)
	validation 15%	33	35	14	26
	test 15%	33	35	14	26

The size of data sets (before and after augmentation) in recognizing different realizations of the sounds /s/ and /ʃ/

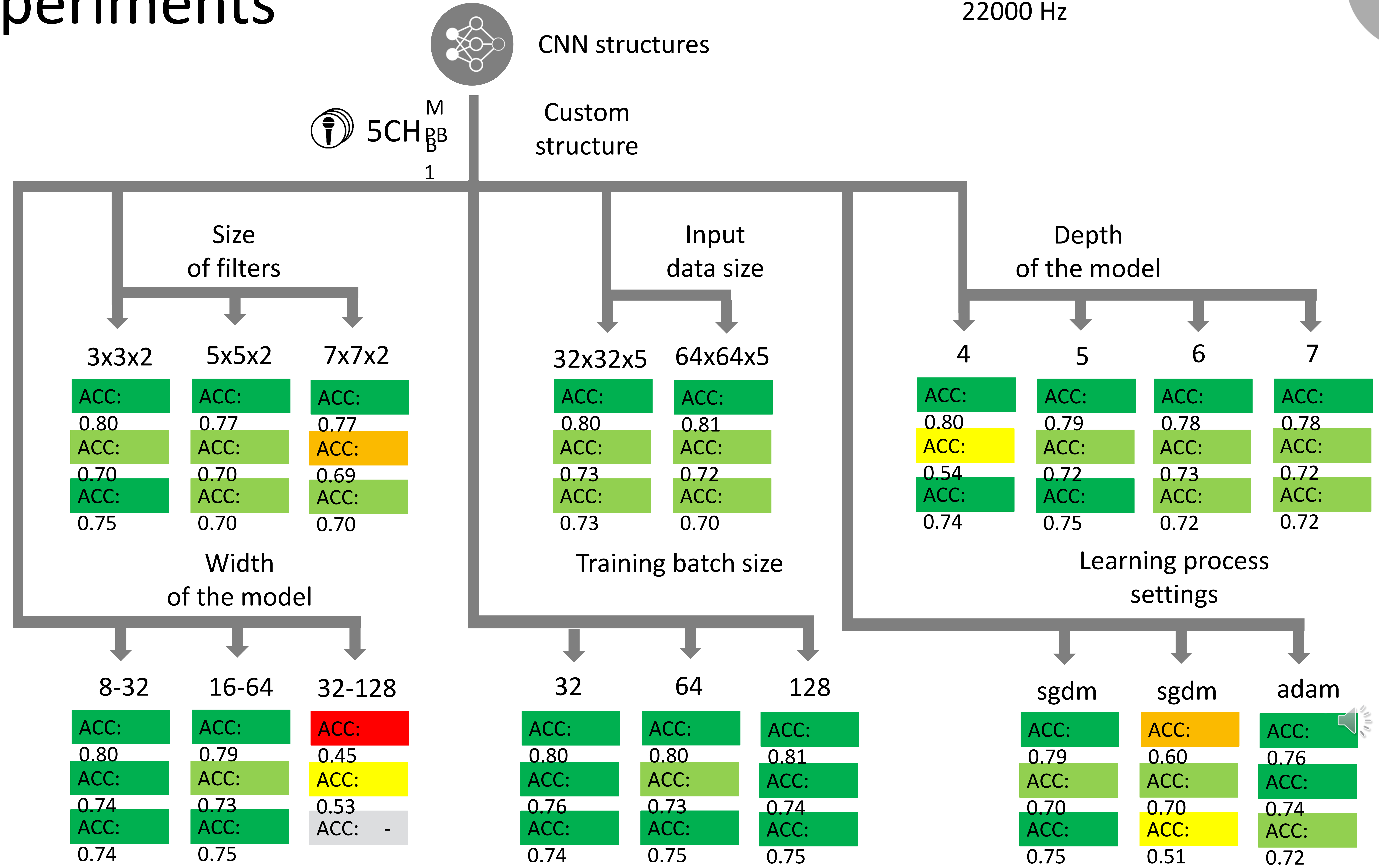
Experiments

linear scale filter range: 1000-22000 Hz



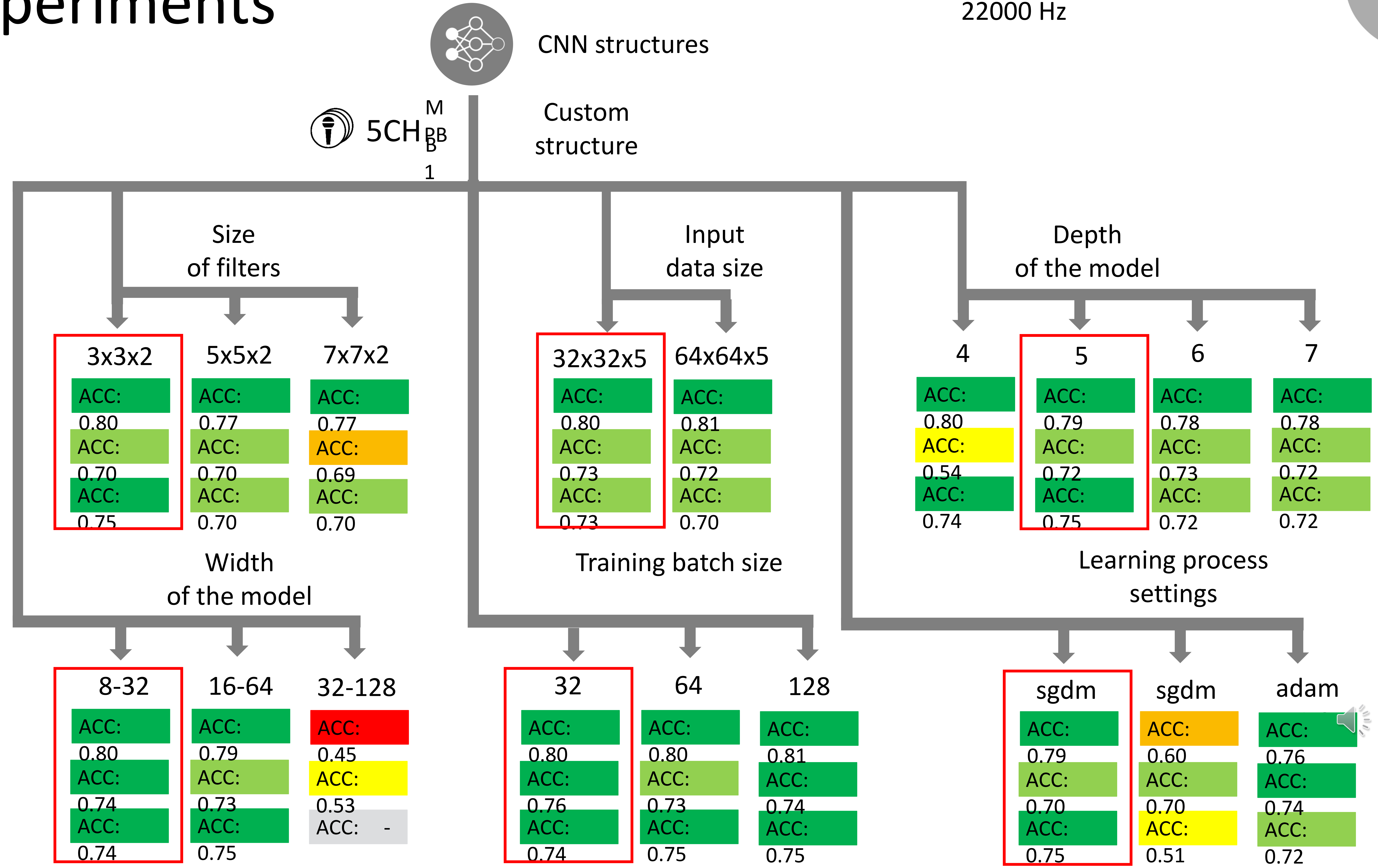
Experiments

linear scale filter range: 1000-22000 Hz

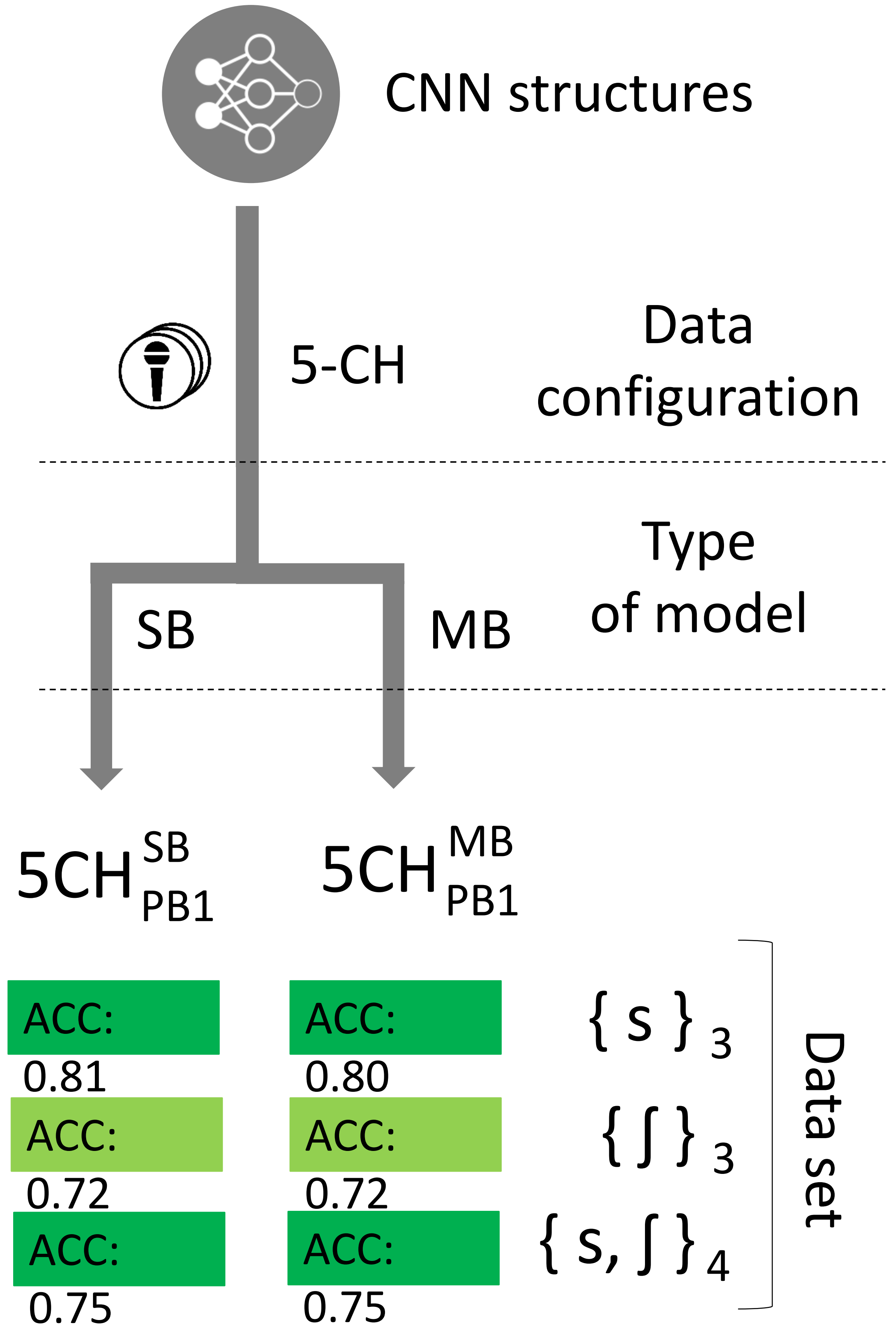


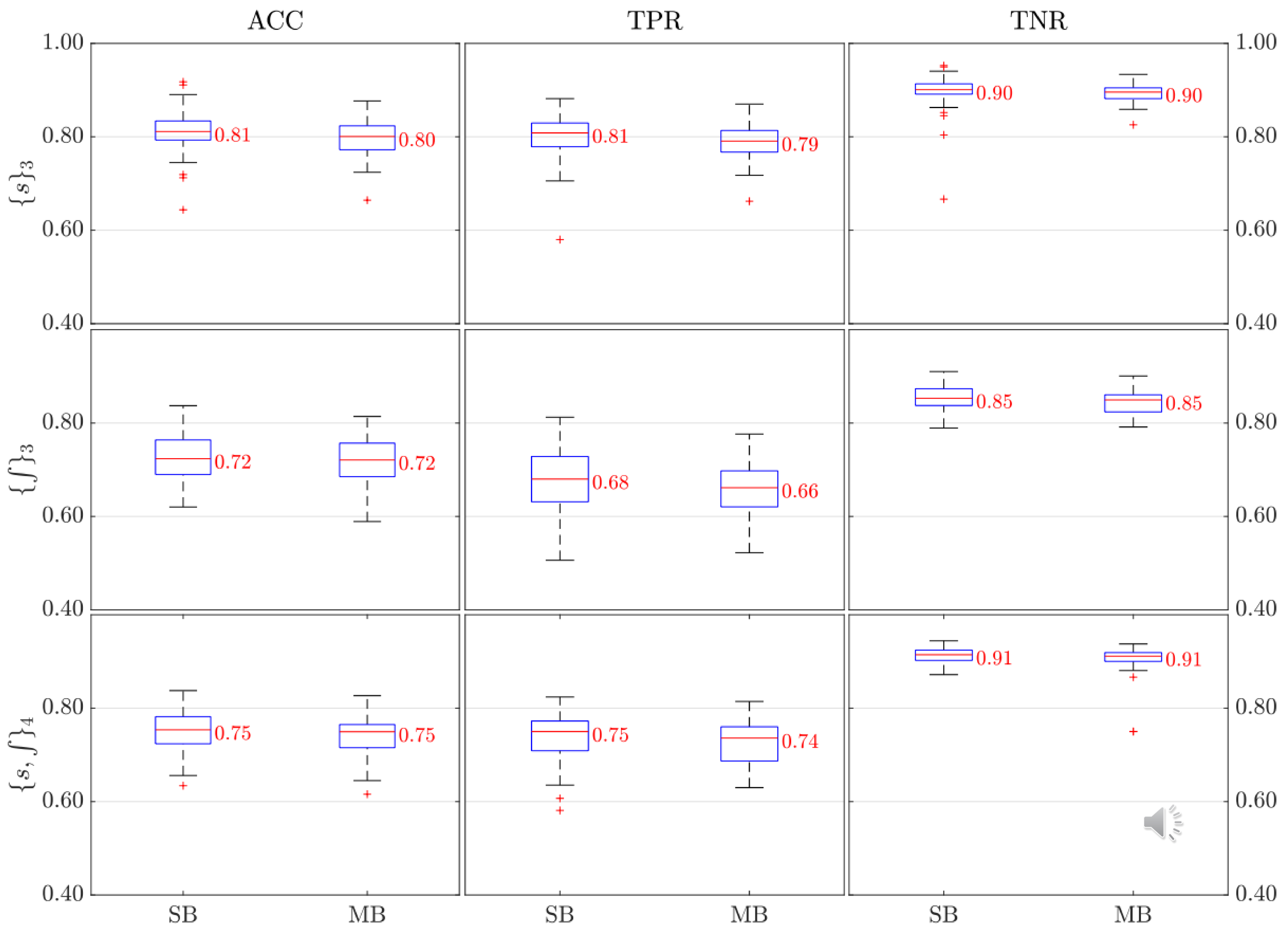
Experiments

linear scale filter range: 1000-22000 Hz



Experiments





Summary

- Design, construction and verification of a measuring device.
- Development of a methodology for spatial processing of the speech signal dedicated to dentalized sounds.
- Development of the measuring device along with the expansion of the speech database, extension of the analysis to include additional dentalized sounds.

Reference

- **Automated detection of sigmatism using deep learning applied to multichannel speech signal**
  Krecichwost, M., Mocko, N., Badura, P.
  Biomedical Signal Processing and Control, 68, 102612. Elsevier, 2021.

- **Hybridization of Acoustic and Visual Features of Polish Sibilants Produced by Children for Computer Speech Diagnosis**
  Sage, A., Miodonska, Z., Kręcichwost, M., Badura, P.
  Sensors, 24(16), 5360. Multidisciplinary Digital Publishing Institute, 2024.

- **Multichannel acoustic analysis of phoneme/s/mispronunciation for lateral sigmatism detection**
  Krecichwost, M., Miodonska, Z., Badura, P., Trzaskalik, J., Mocko, N.
  Biocybernetics and Biomedical Engineering, 39(1), 246-255. Elsevier, 2019.

- **Multichannel Speech Acquisition and Analysis for Computer-Aided Sigmatism Diagnosis in Children**
  Krecichwost, M., Miodonska, Z., Trzaskalik, J., Badura, P.
  IEEE Access. IEEE, 2020.