A Computational Approach to Analysis and Detection of Singing Techniques

January 29th, 2024 Ph.D. Defense

Yuya Yamamoto

Informatics Program, University of Tsukuba

Supervisers:
Nobutaka Suzuki
Hiroko Terasawa
Hiroyoshi Ito

Committee:
Nobutaka Suzuki
Hiroko Terasawa
Atsushi Toshimori
Shuichi Moritsugu
Juhan Nam

♬Can you listen to the music??

Outline

- **Introduction**
  - Background
  - Research Aims and Problem Statements
- **Research works**
  - Ch. 3: Exploration of Singing Techniques
  - Ch. 4: Singing Technique Analysis on Actual Vocal Performances
  - Ch. 5: Characteristics-aware Modeling for Singing Technique Classification
  - Ch. 6: Singing Technique Detection from Real-world Vocal Tracks
- **Conclusion**

Introduction

Combined Chapter 1 and Chapter 2 in the presentation
Chapter 1: Motivation, problems, and aims of the thesis
Chapter 2: Related works

Research area: Singing voices in music

- Singing voice is one of the most important parts in music

- Singing, Listening, Creating etc. are fundamental cultural and artistic activities for human beings

→ Many research works have been done to clarify or to engineer about singing voice

Singing technique is one of the ways to embody the expression

🎵Bohemian Rhapsody/ Queen

Techniques: Vibrato (yellow), Portamento (pink), Falsetto (blue), Raspy voice (White)

Sore Demo Shitai / Ken Hirai

♫ Playing…

Lyrics

A Na Ta Ha Ke-**Shi** **Te**

Tsu Ka Wa **Na** I **No** **Yo**

A Ta Shi No Su Ki Na

Bo **Di** So - Pu **Wo**

- **Vocal fry**: producing pulsive sound
- **Bend**: short going-around pitch bending
- **Vibrato**: periodic pitch modulation
- **Scooping**: short ascending pitch bending

Shin Ji Dai / Ado

🎵 Playing…

Lyrics

Shin Ji Da I **Ha**

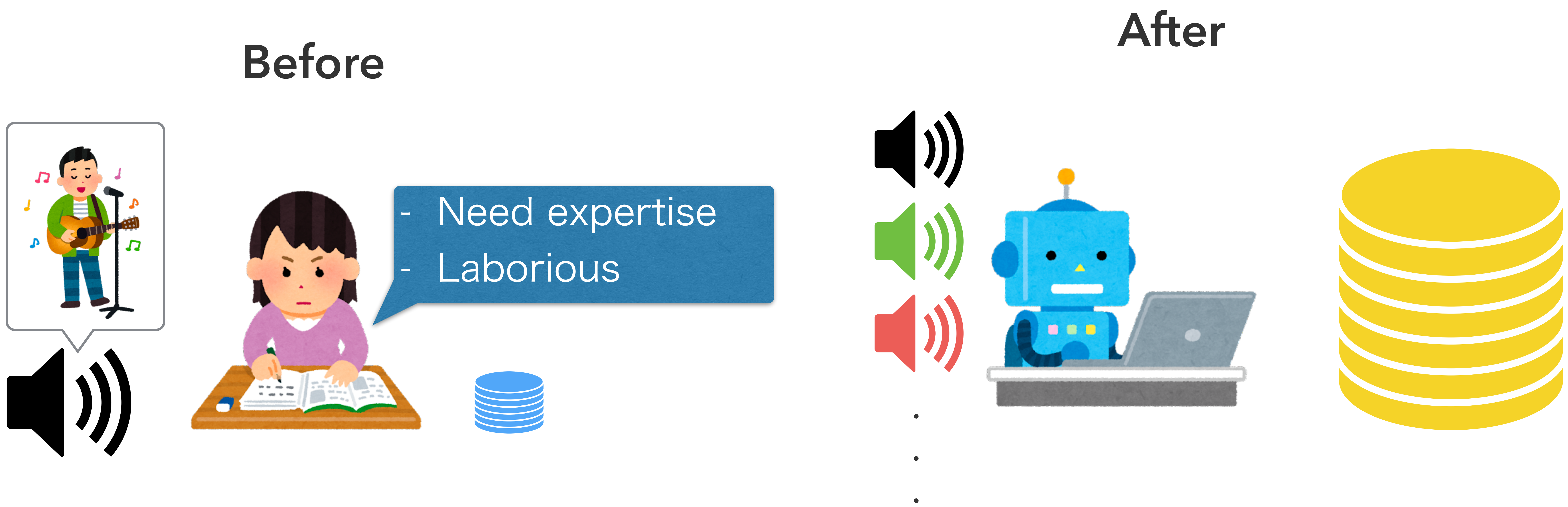
Ko No Mi **Ra** I Da

S(e) Kai Ju **U** Ze M **Bu**

Ka E Te Shi **Ma E Ba**…

- **Falsetto**: different tone on higher pitch
- **Vibrato**: periodic pitch modulation
- **Scooping**: short ascending pitch bending
- **Hiccup**: short falsetto & pitch jump
- _, (): Enhancing/Omitting phonemes

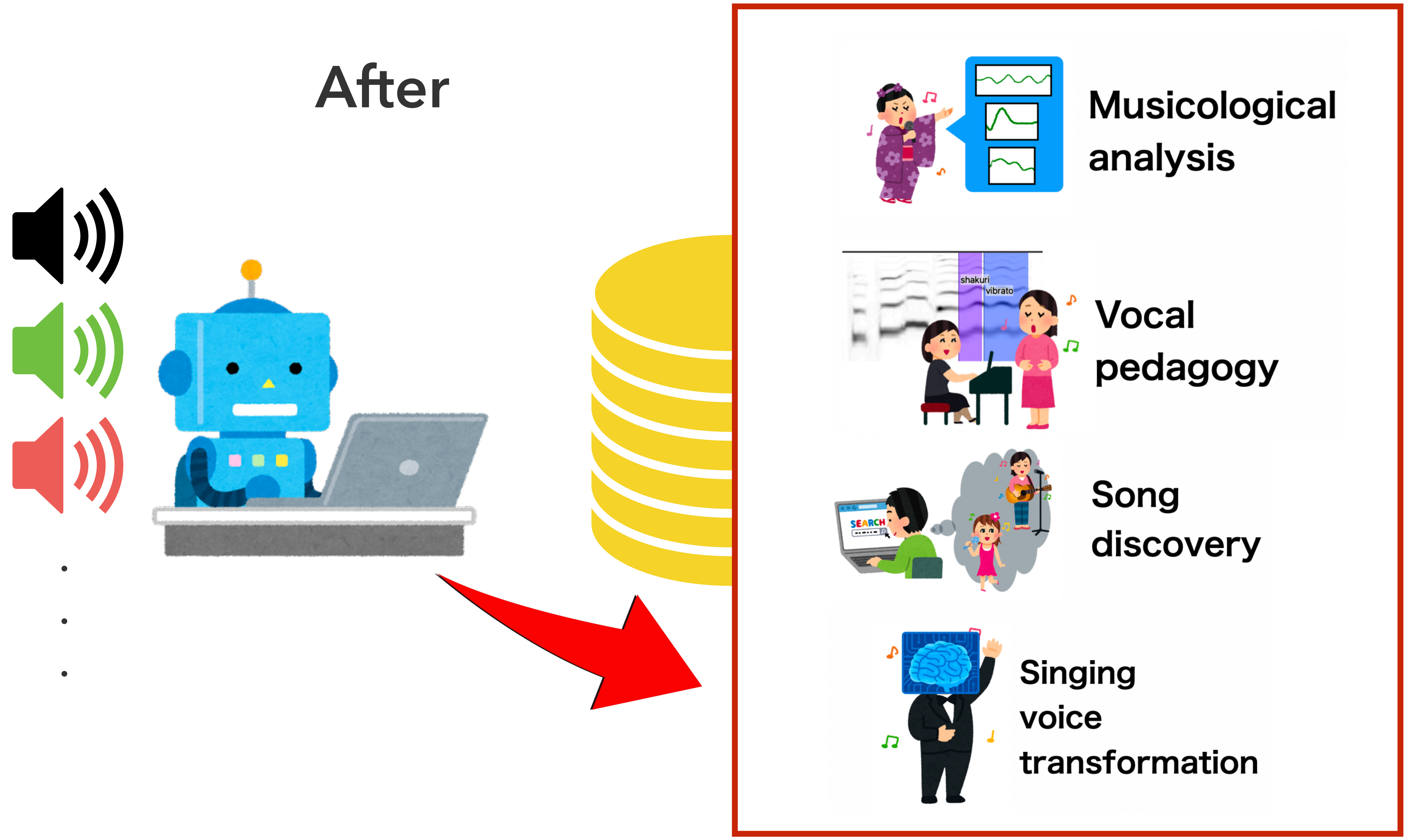
Establish "Computational" framework for singing technique analysis



- Advantages
  - **1. Accelerates the singing technique analysis by automation**

Aim of the research

Establish “Computational” framework for singing technique analysis



- Advantages
  - 1. Accelerates the singing technique analysis by automation
  - 2. Combines to various computational applications

(e.g., analysis, pedagogy, discovery, creation, etc.)

Related works ① Nature of singing techniques

- **Manually analysis of singing styles to clarify the characteristics**
  - Recording & Acoustic parameter analysis (e.g., Vibrato parameter [Seashore 37])
  - Qualitative song reading (e.g., “Kobushi” in Ken Hirai [Nakazato 09])

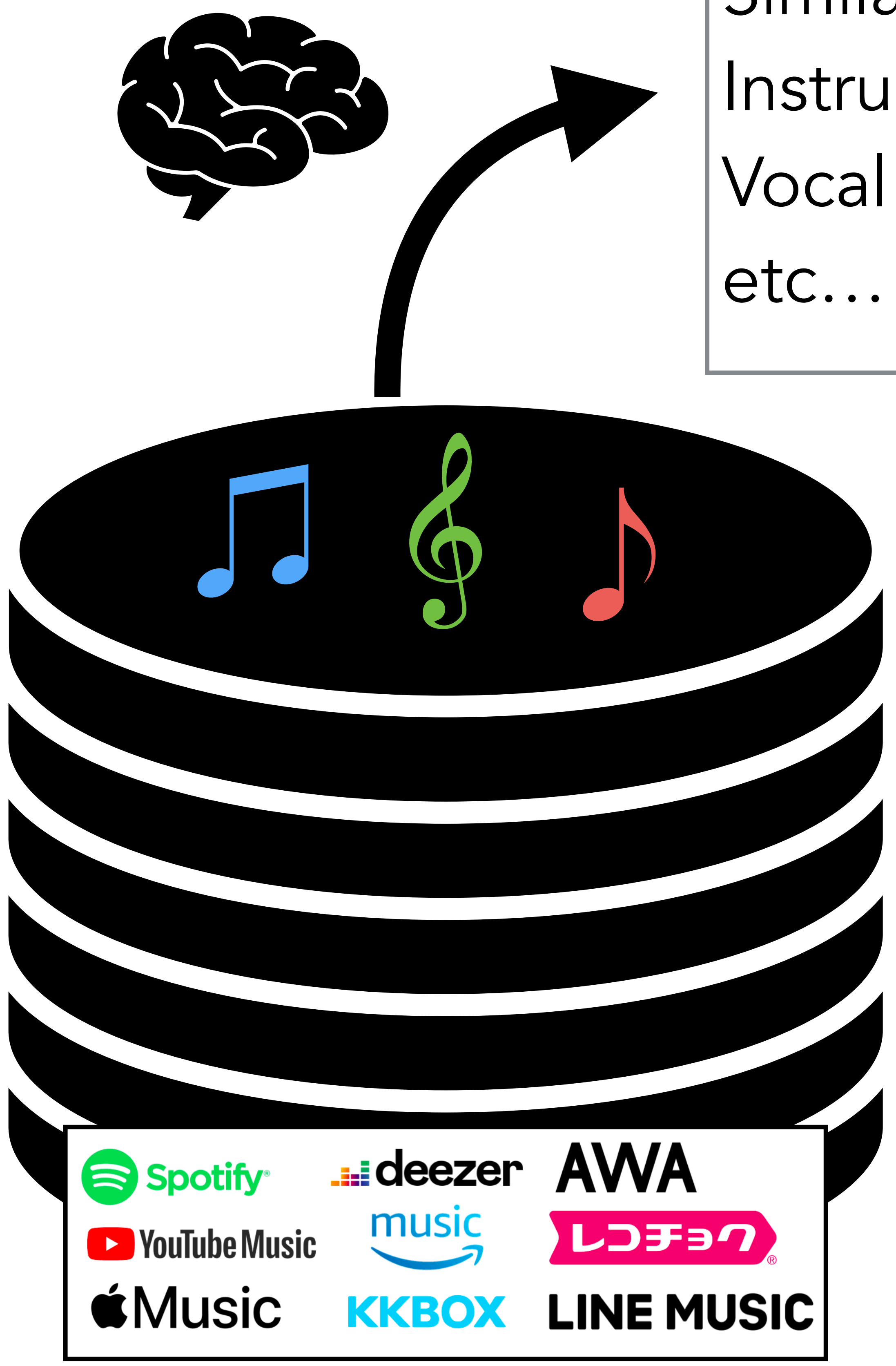
- **Valuable, yet still less applicable in a real-world scenario**
  - Lacks overview and objectivity
  - Needs expertise and time-consuming
  - Few samples and limited scope

Research field: Music information retrieval (MIR)

Extract information

Genre
Mood
Similarity between others
Instruments
Vocal style
etc…

MIR methods



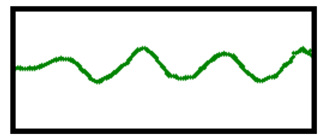
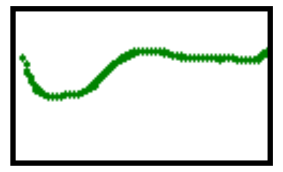
Existing music data

- **Techniques analyzing music automatically**
  - Enrich human's musical activity by mining information from music data
  - Recently enhanced by data-driven approach
- **Singing voice analysis is an important topic in MIR**
  - Singing voice is center of music contents
  - Relates to listener's music preference [Demetrou18]

→MIR can realize the automation!

[Demetrou18] A. Demetrou et al. VOCALS IN MUSIC MATTER: THE RELEVANCE OF VOCALS IN THE MINDS OF LISTENERS, ISMIR 2018

Related works ② Automatic vocal analysis in MIR

	Song	Vocalist	Expression
component			 Vibrato at each "go"  Glissando at first "let"
role in performance	**What?**	**Who?**	**How?**
MIR tasks/ research field	Singing note- / lyric- transcription	Singer identification	Singing technique analysis
Conventional works	[Mauch 15], [Nishikimi 17], [Gupta 20], [Hsu 21], [Wang 21], [Demirel 21] and more.	[Fujihara 10], [Kroher 14], [Wang 18] [Nachmani 19], [Lee 19], [Hsieh 20] and more.	- Indian raga analysis, 4 pitch techniques [Miryala 13] - Vocal style transfer, 4 pitch techniques [Ikemiya 14] - Track-wise identification, 10 various techniques [Wilkins 18] - Heavy metal scream detection, 3 different scream [Kalbag 22]

Related works ③ Datasets of singing technique

Lacking in-song, various techniques, and with timestamp annotation

Dataset	singing songs	annotation with timestamp	monophonic (a capella)	genre/purpose	kinds	numbers
Phonation Modes [Proutskova 13]			✓	Classic/ Phonation mode classification	4	763
VocalSet [Wilkins 18]	(Simple phrase e.g., scale, arrpegio)		✓	Classic and popular/ Singer technique classification	10	3560
KVT Dataset [Kim 20]	✓			K-POP/ Singing voice tagging	7	446
MVD Dataset [Kalbag 22]	✓	✓		Heavy metal/ Scream detection	3	57
SVQTD [Xu 22]	✓			Classic/singing attribute classification	7	4000

[Proutskova 13] Breathy, Resonant, Pressed – Automatic Detection of Phonation Mode from Audio Recordings of Singing, P. Proutskova et al., Journal of New Music Research, 2013
[Wilkins 18] VocalSet: A Singing Voice Dataset, J.Wilkins et al., ISMIR 2018
[Kim 20] Semantic Tagging of Singing Voices in Popular Music Recordings, K. L Kim et al., TASLP 2020.
[Kalbag 22] Scream Detection in Heavy Metal Music, V. Kalbag et al. SMC 2022
[Xu 22] Paralinguistic singing attribute recognition using supervised machine learning for describing the classical tenor solo singing voice in vocal pedagogy, Y. Xu et al. EURASIP JASM 2022.

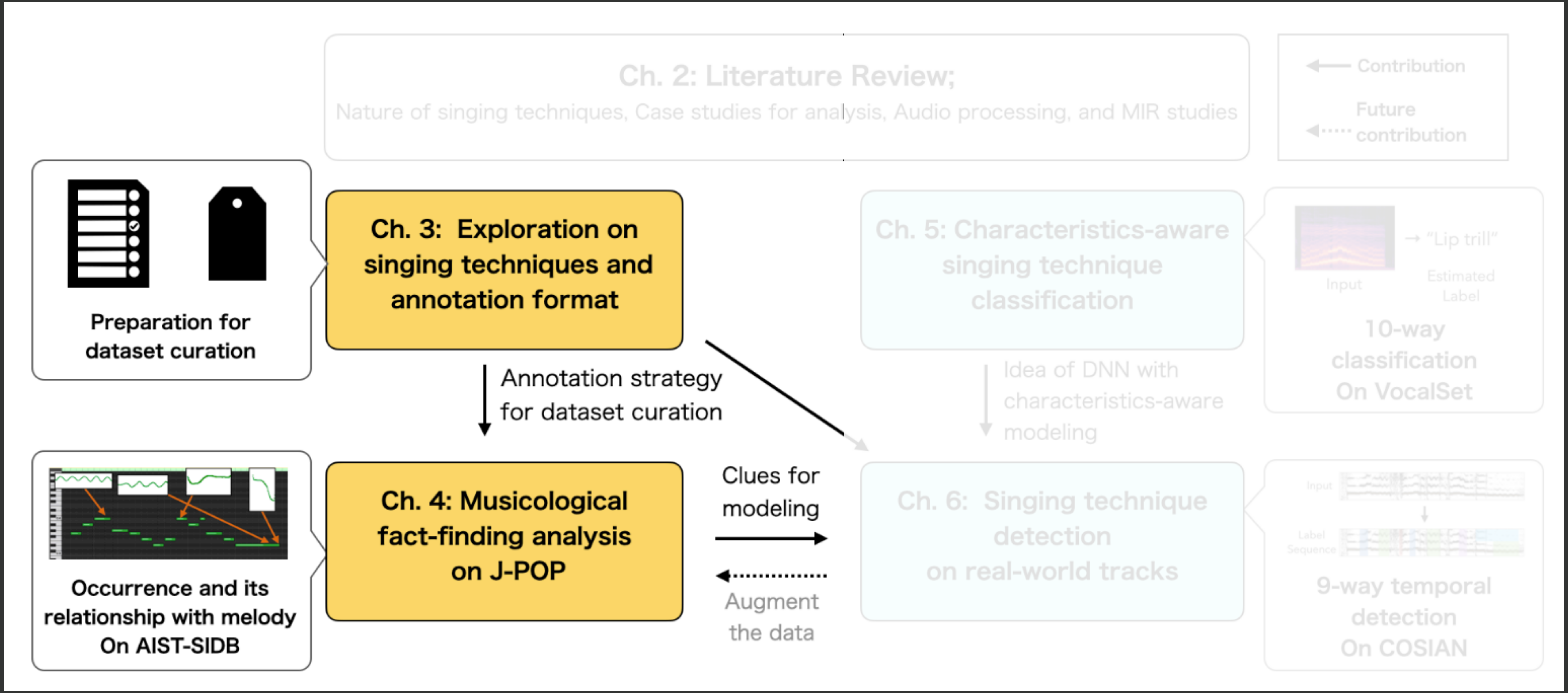
The lack on framework; dataset and computational methodology

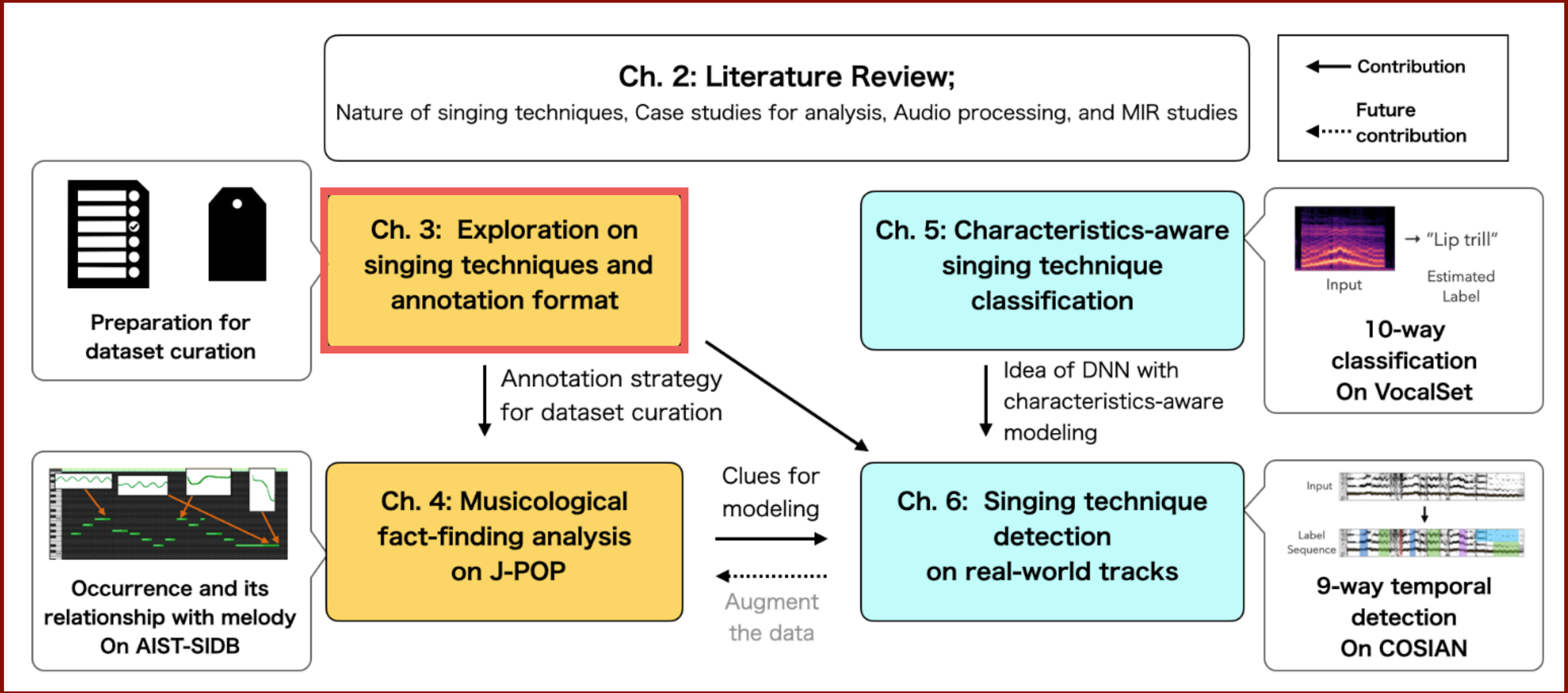
- **1. Absence of data, annotation and its characteristics**
  - What singing techniques should be annotated?
  - Need datasets, but how to annotate?
  - What is their specific characteristics?

- **2. Less established identification methods**
  - How to design the automatic model?
  - Can we detect techniques from real-world singing tracks?

Proposal: Computational foundations for singing technique analysis

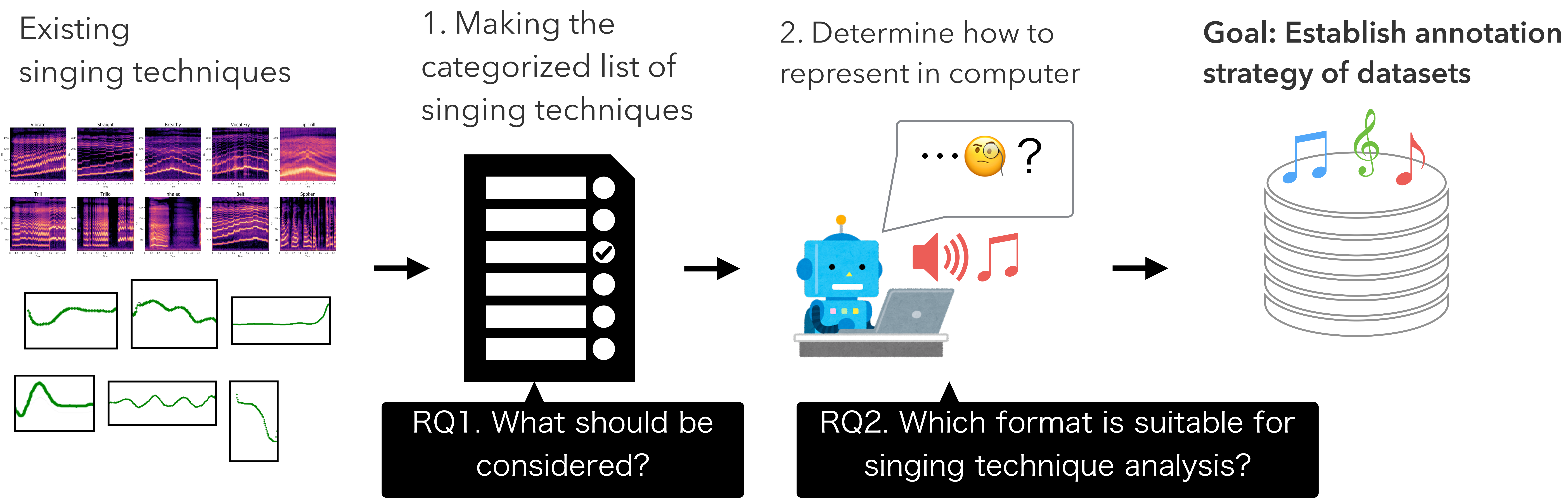
Part I. Data exploration





Chapter 3
Exploration of Singing Techniques for Dataset Creation

Exploration of existing singing techniques to build datasets

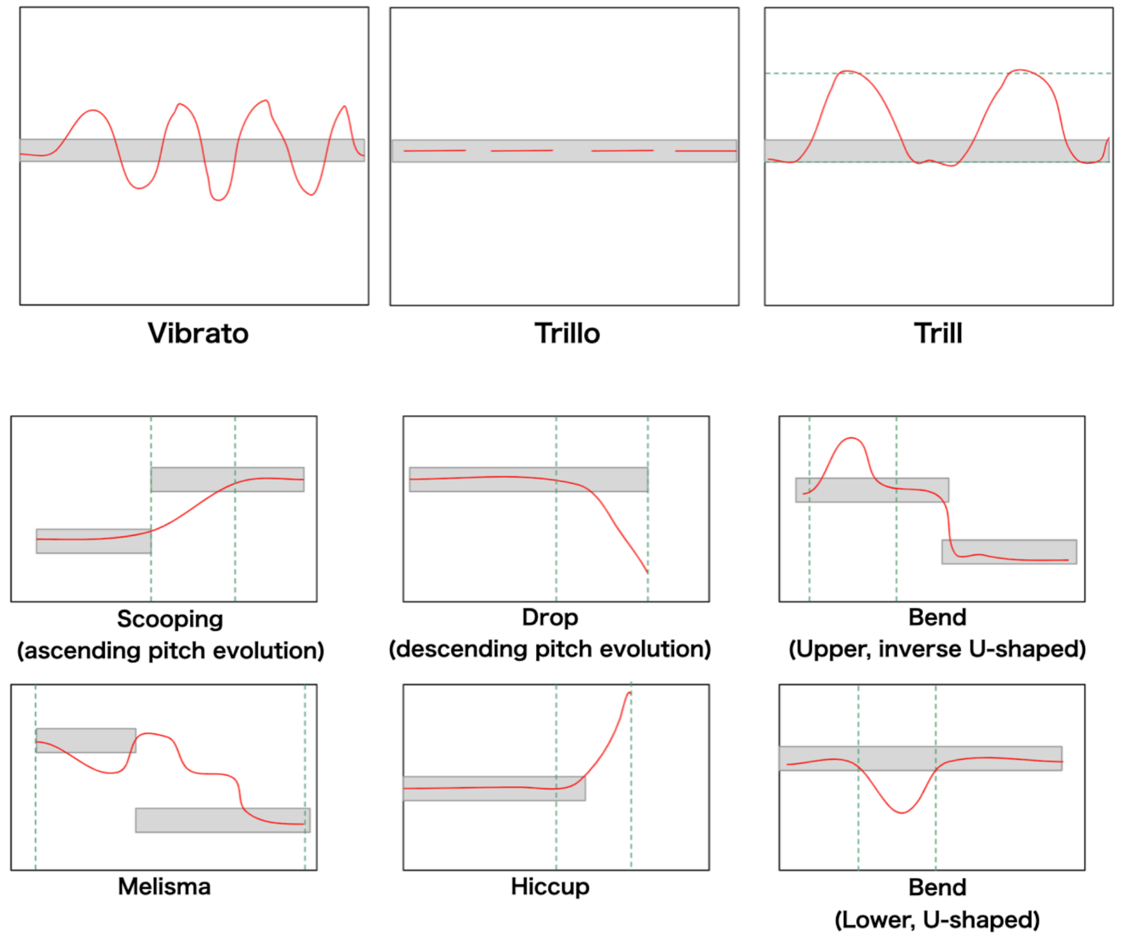


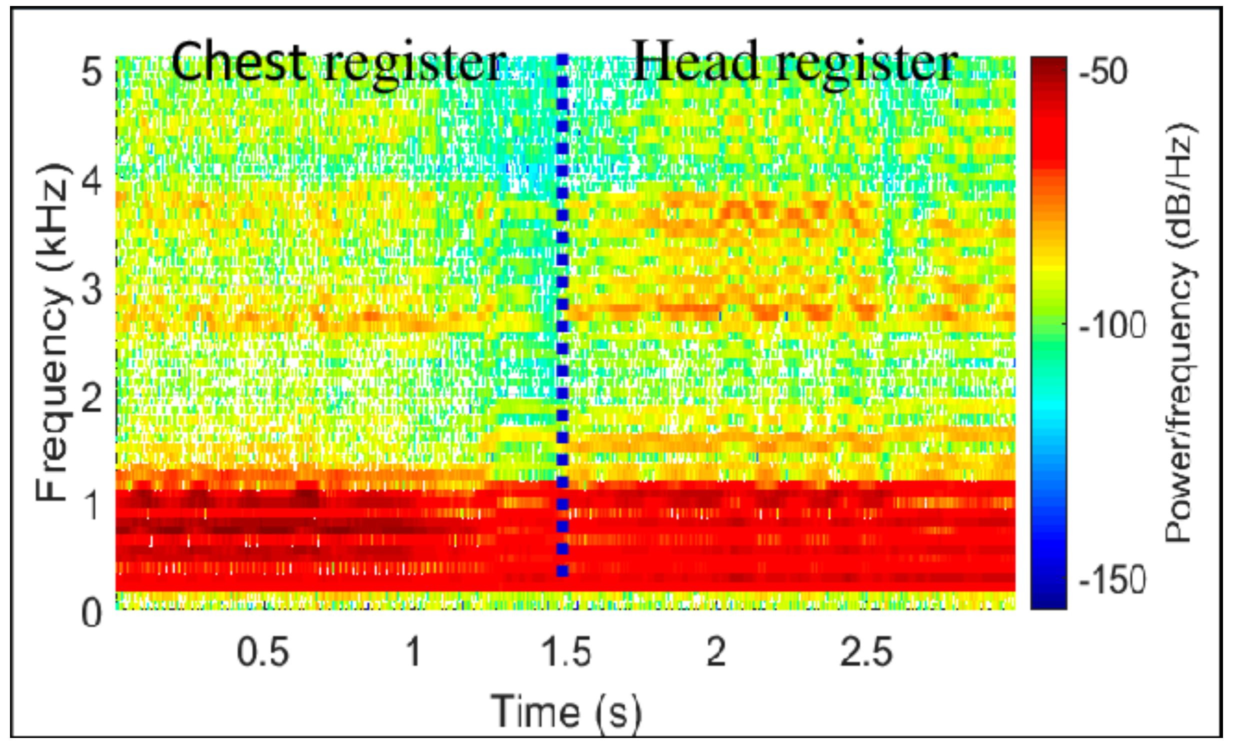
- **Contributions:**
  - 1. Listed up named singing techniques from various musical aspects (e.g., pitch, timbre, etc.) -> RQ1
  - 2. Adopt region label (name+ start & end time) for annotation -> RQ2

Literature review

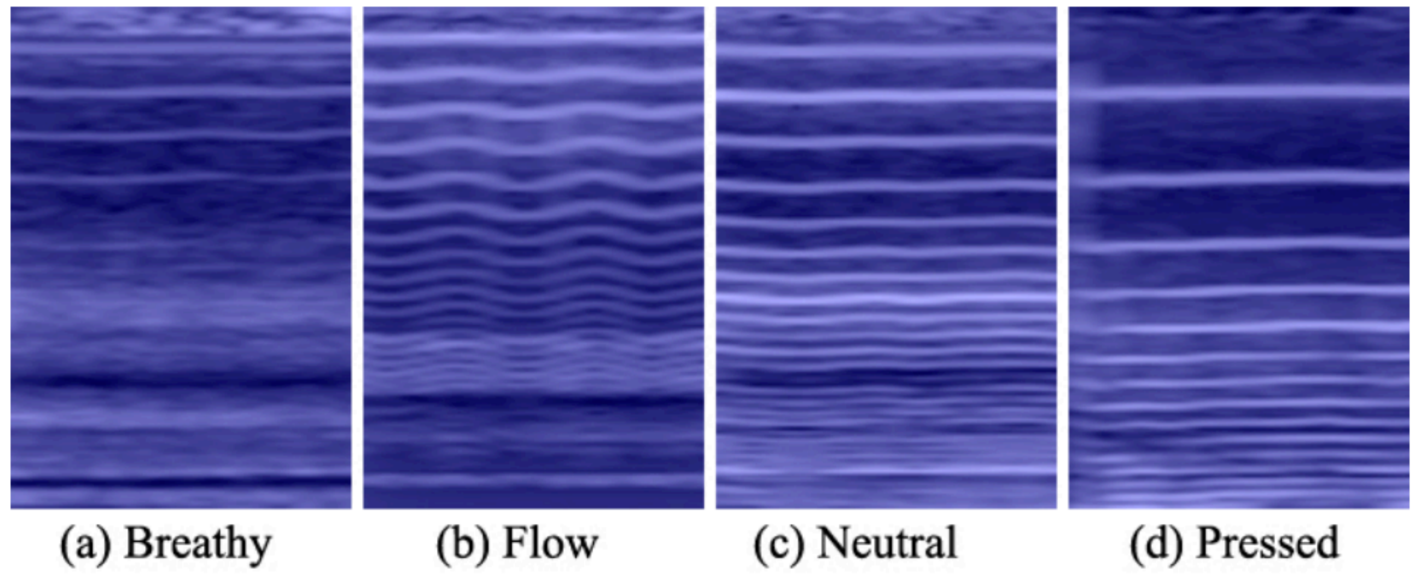
Listed named techniques by literature survey both on academic and non-academic

- Explored on various named techniques in the world
- Categorize coarsely by what is the fluctuated components
  - Pitch: modulation, portamento (continuously changing), bending.
  - Timbre: register, phonation mode, extreme effects, etc.
  - Others





R. Elbarougy, Acoustic Analysis for Chest-to-Head Register Transition in Singing Voice International Journal of Computer Applications 177(10):11-16, 2019



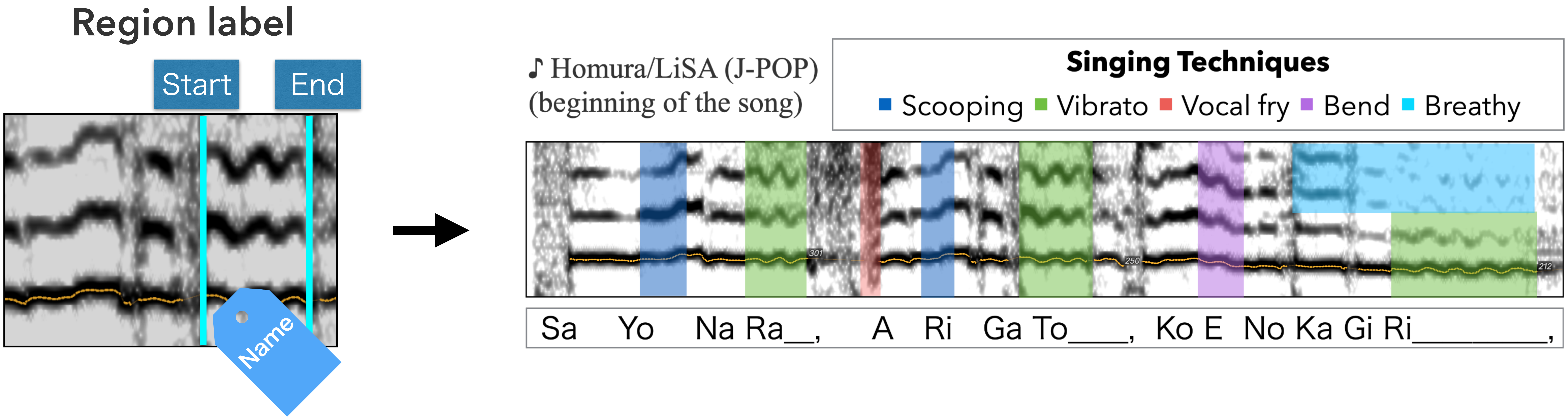
X. Sun et al. RESIDUAL ATTENTION BASED NETWORK FOR AUTOMATIC CLASSIFICATION OF PHONATION MODES, arxiv 2021.

Candidates of singing techniques

For more detail:
Refer to the thesis!

Annotation strategy

- 1. Represent by region labels (name+ start&end time boundaries)
- 2. Based on observable singing techniques in data

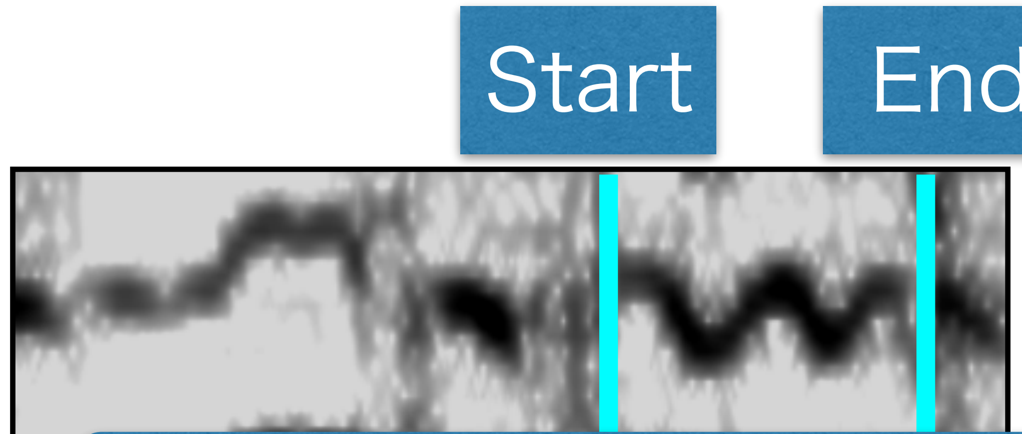


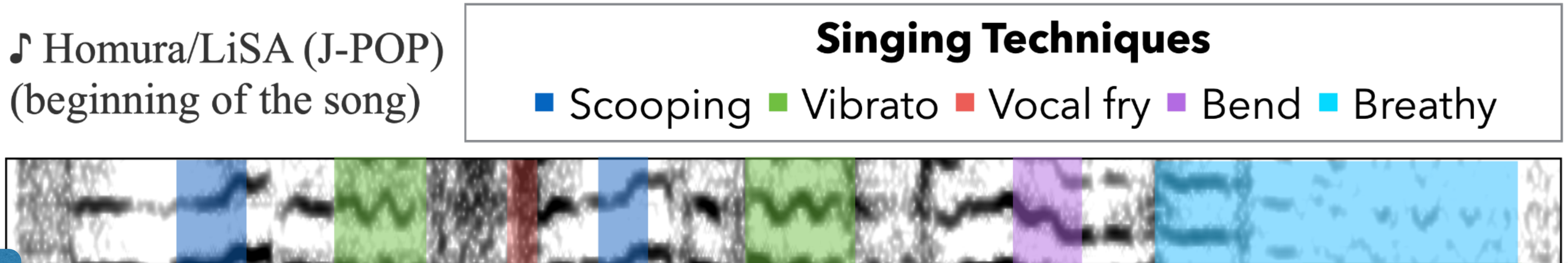
- **Advantages: ① Can represents local phenomena, ② Intuitive, ③ Note-information-free**

Annotation strategy

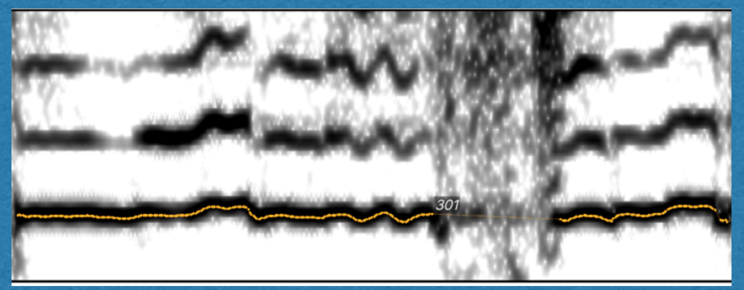
- 1. Represent by region labels (name+ start&end time boundaries)
- 2. Based on observable singing techniques in data

Region label





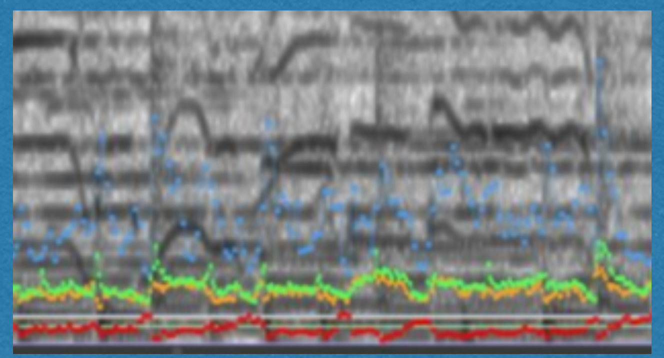
VS. Global tag



-> vibrato, Scooping

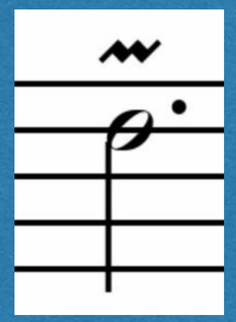
Can not represents location

VS. Numeric parameters



Less intuitive for humans

VS. Note-wise label

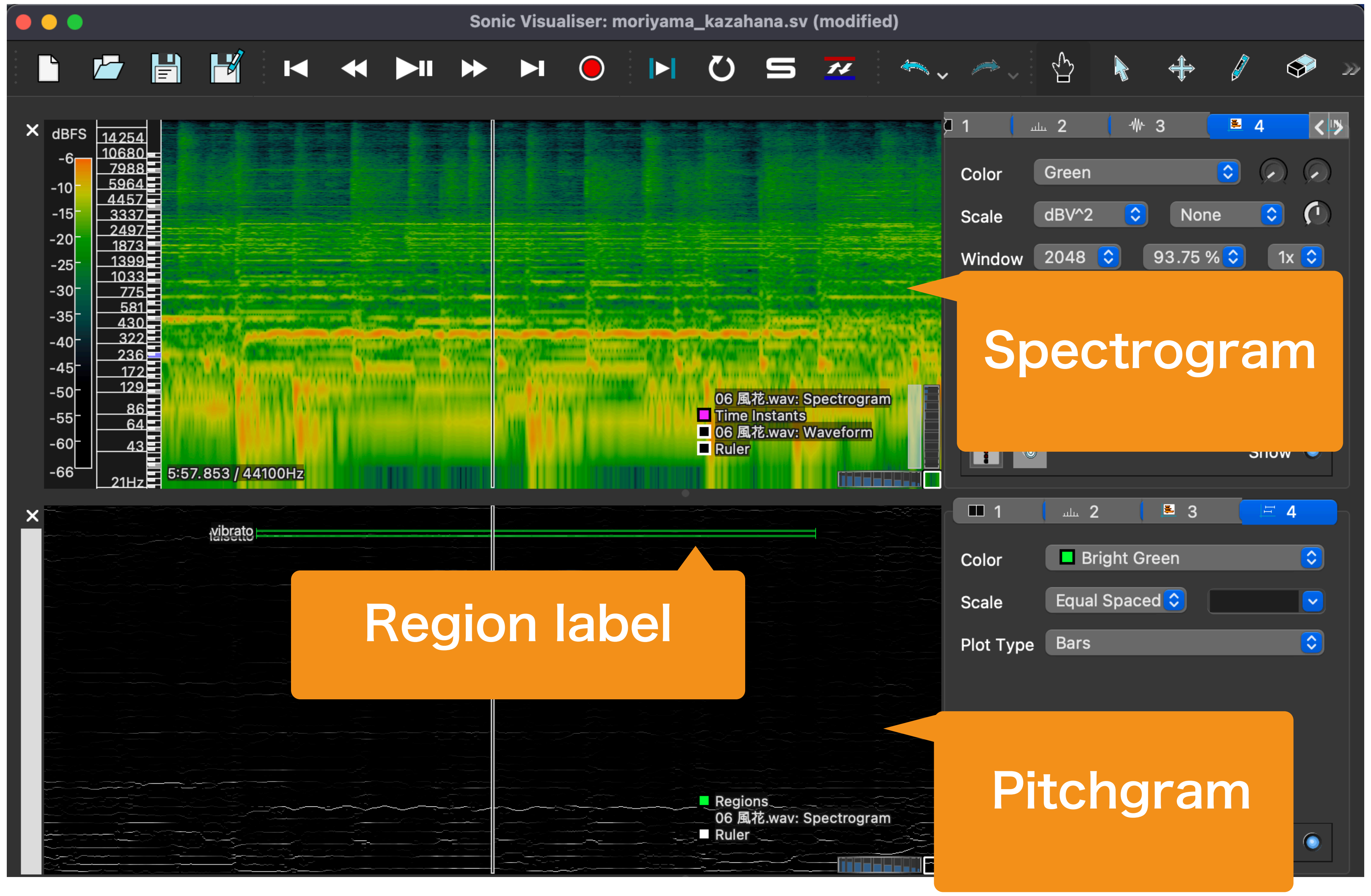


Needs additional note annotation

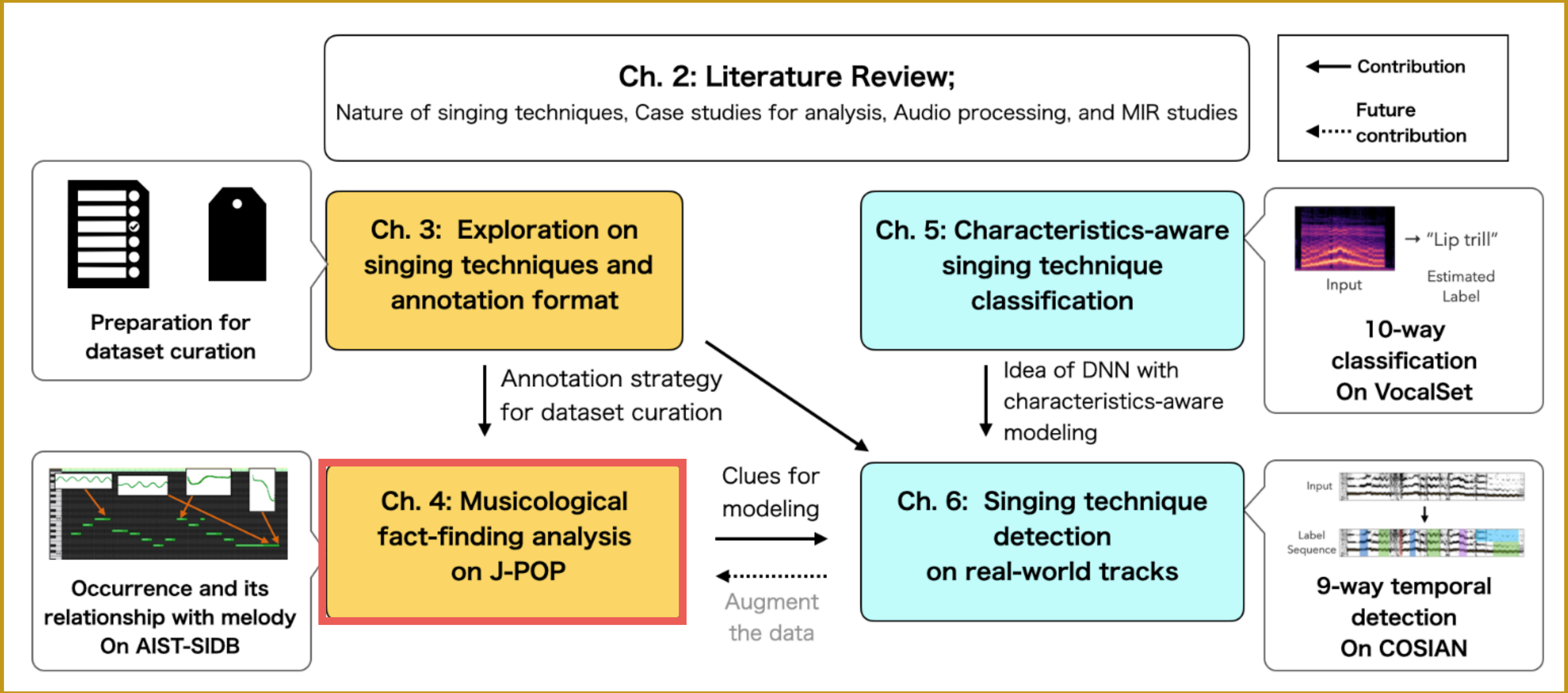
- **Advantages: ① Can represents local phenomena, ② Intuitive, ③ Note-information-free**

Annotation process

- **Manually annotated**
  - **Annotator**: author
    - amateur (no academic degree on music)
    - 9 years of popular vocal, 4 years of chorus (tenor),
    - has relative pitch
  - **Software**: Sonic visualiser [Cannam 10]
    - visualizing spectrogram and pitchgram
    - both aid of visual & audio feedback
    - set region label on pitchgram



[Cannam 10] Cannam et al. Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. ACMMM 2010

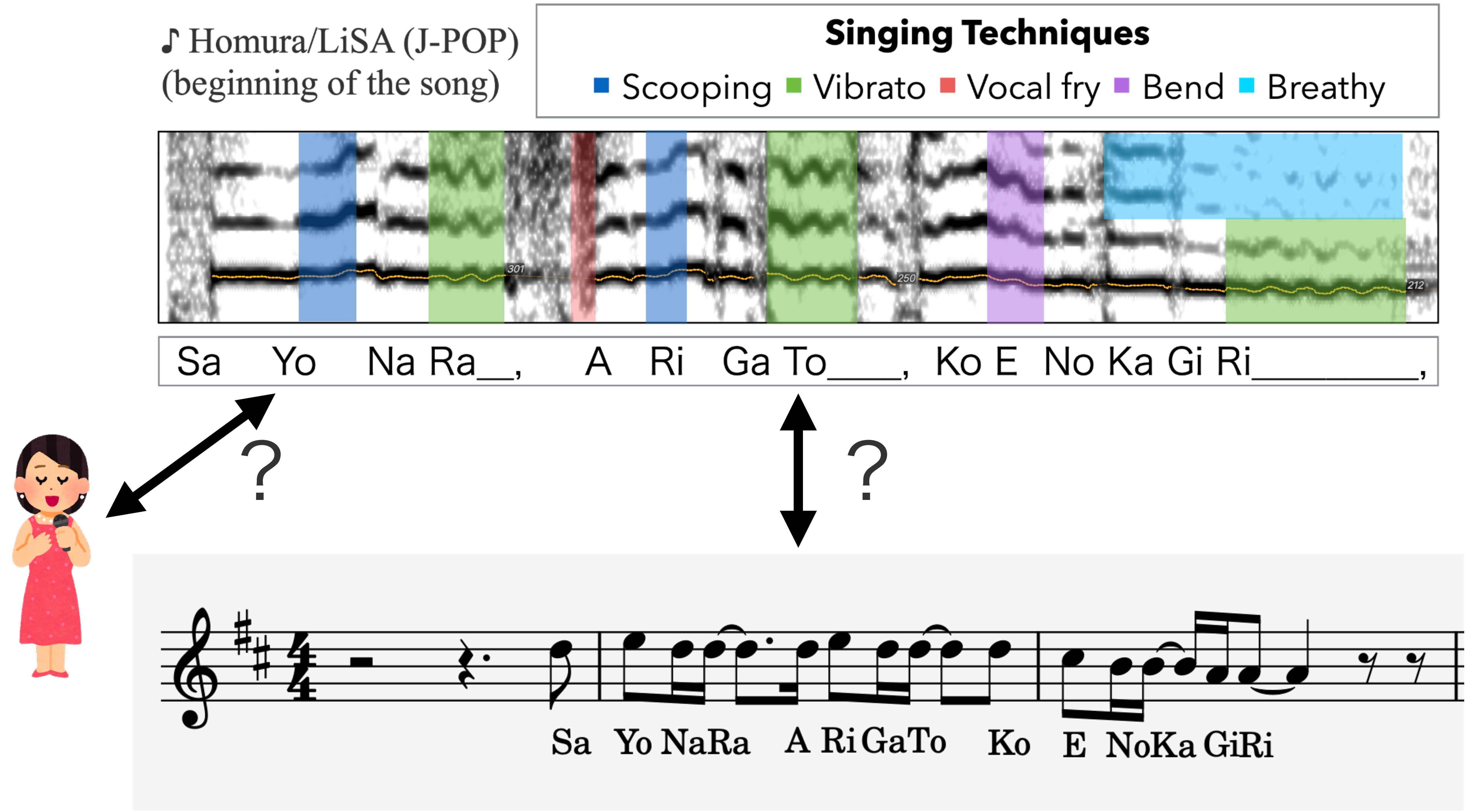


Chapter 4
Musicological Analysis of Singing Techniques with Correspondence to Musical Score on Imitative Singing

Overview of Chapter 4

Investigate the relationship between singing technique and song

- **1. Occurrence frequency**
  (RQ: How many on each singer?)
- **2. Vibrato parameter**
  (RQ: How to produce, where?)
- **3. Occurrence locations**
  (RQ: How many where in the song?)
- **Goal: Better understanding singing techniques**



- **Contributions**
  - New discoveries about tendency of singing technique appearance
  - Enhanced utility of singing techniques annotation and their analysis by showing the relationship with singer and song

Data: AIST Singing Imitation Database (AIST-SIDB)

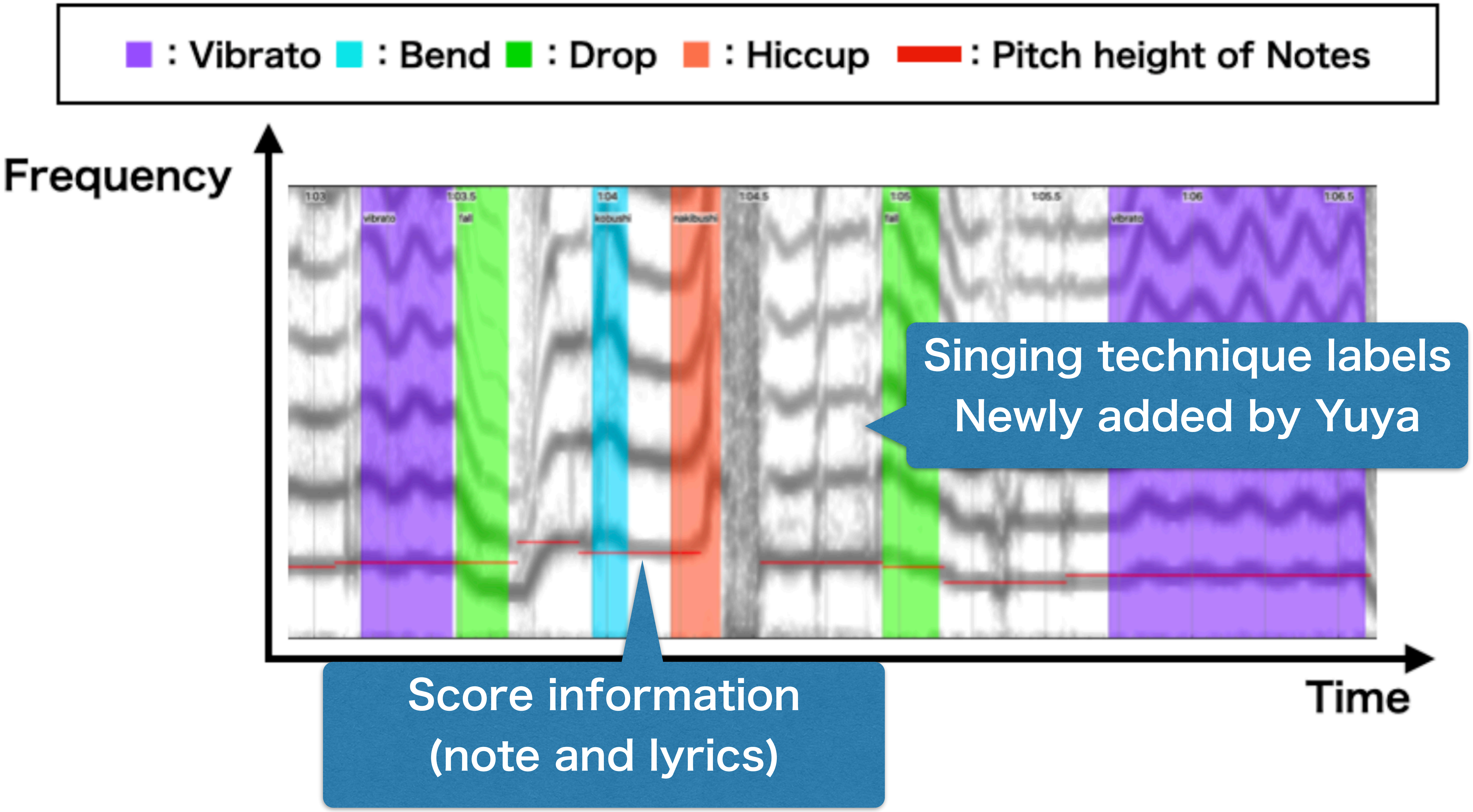
- **Imitation of J-POP famous singers**
  - A Cappella, studio quality
  - Original：24 (12 for each gender)
  - Imitator: professional singer (7 F/M)
  - 48 tracks（two imitator per song)
  - **Private database, possessed by AIST**

- **Examples**

Keisuke Kuwata/ Katte ni Sinbat
Imitator1 **Imitator2**
♪Su Na Ma Ji Ri No Chi Ga Sa Ki
Hi To Mo Na Mi Mo Ki E Te…

	Original singer	Song name	Gender	Imitator1	Imitator2
	玉置浩二（Tamaki）	出逢い	M	M03	M04
	小田和正（oda）	キラキラ	M	M02	M06
	Gackt（gackt）	ありったけの愛で	M	M01	M05
★	桑田佳祐（sazan）	勝手にシンドバット	M	M06	M07
	チバユウスケ（skapara）	カナリヤ鳴く空	M	M05	M01
	西川貴教（tmr）	Heat Capacity	M	M05	M01
	hyde（larcenciel）	Lies and Truth	M	M03	M04
	平井堅（hirai）	瞳を閉じて	M	M01	M05
	福山雅治（fukuyama）	桜坂	M	M04	M03
	槇原敬之（makihara）	桃	M	M04	M03
	森山直太朗（moriyama）	さくら（独唱）	M	M02	M06
	山崎まさよし（yamazaki）	未完成	M	M06	M07
★	aiko（aiko）	ボーイフレンド	F	F01	F06
	絢香（ayaka）	三日月	F	F05	F02
	宇多田ヒカル（utada）	Can You Keep A Secret?	F	F03	F04
	鬼束ちひろ（onitsuka）	月光	F	F03	F04
	倖田來未（koda）	夢のうた	F	F06	F01
	小柳ゆき（koyanagi）	愛情	F	F04	F07
	chara（chara）	大切をきずくもの	F	F02	F05
	浜崎あゆみ（hamasaki）	seasons	F	F06	F01
	一青窈（hitotoyo）	ハナミズキ	F	F05	F02
	平原綾香（hirahara）	明日	F	F03	F04
	松浦亜弥（matsuura）	♡ 桃色片思い ♡	F	F01	F06
	YUKI（jam）	motto	F	F02 (key-1)	F05

Annotation for AIST-SIDB



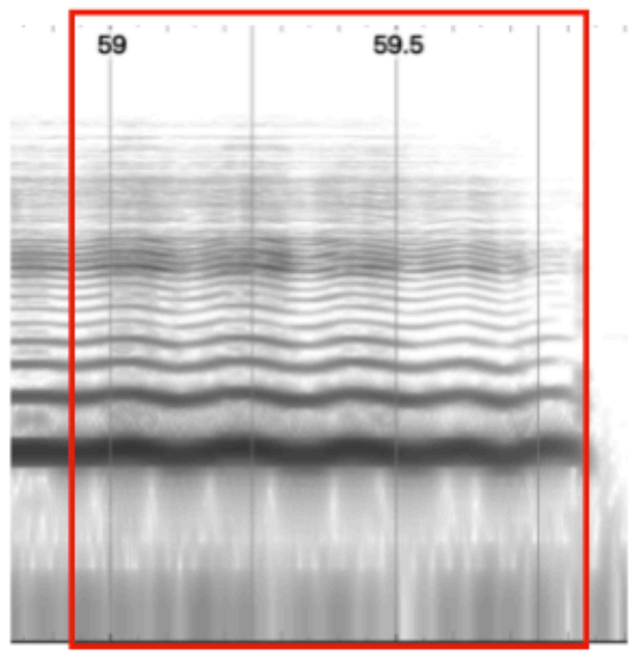
Annotated techniques

- Annotated 13 singing techniques
- Based on the survey in Chapter 3 and observation of data

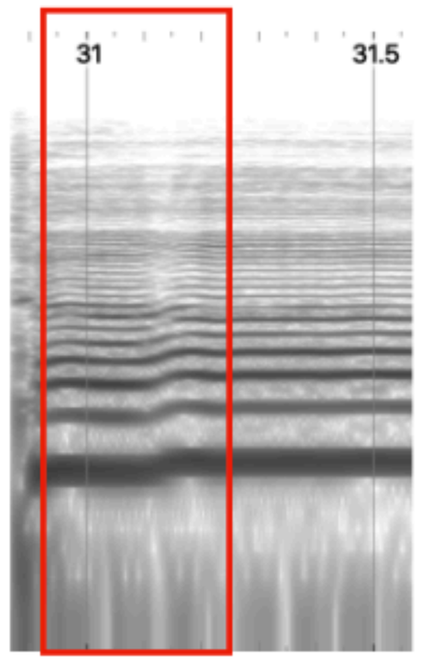
Table 4.2: Definitions, synonyms, and comparable techniques for singing techniques.

Techniques	What is modulated	How modulates	Discrepancies	Similar techniques
Vibrato	pitch, loudness	Singing with a wavering effect, introducing periodic oscillation	Shake	Tremolo, Trill
Scooping	pitch	Continuously changing pitch upward	Glissando, Portamento, Scoop, Scoop up	
Bend	pitch	Continuously changing pitch in a U-shaped or inverted U-shaped pattern	Bend, Tremolo	
Drop	pitch	Continuously changing pitch downward	Drop, Scoop down	
Hiccup	pitch, timbre	Producing a momentary falsetto or tightened throat singing voice	Cry, Sob, Vocal break	Yodel
Melisma	pitch	Assigning multiple pitches to a single syllable	Fake	
Vocal fry	timbre	Producing a raspy sound	Edge voice, Creaky voice	Growl
Falsetto	timbre	Singing in the falsetto range	Head voice, Falsetto	
Breathy	timbre	Mixing in breathy sounds		
Whisper	timbre	Singing in a whispering manner		
Shout	–	Shouting		
Spoken	–	Singing in a spoken manner		Rap
Tongue trill	–	Using rolled tongue	Tongue roll, Rolled tongue	Lip roll

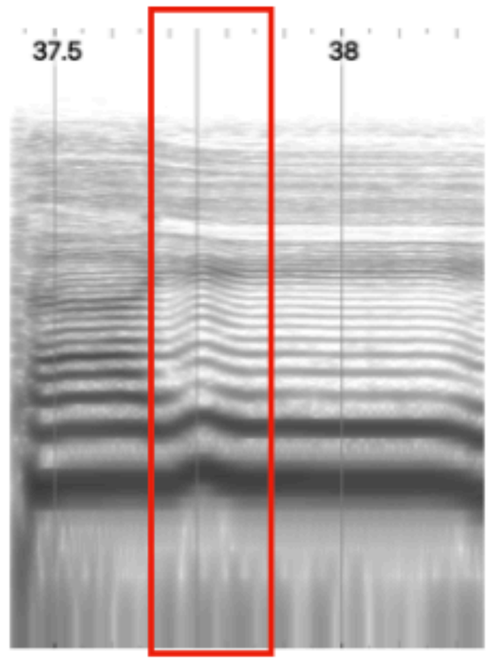
Spectrograms of each technique



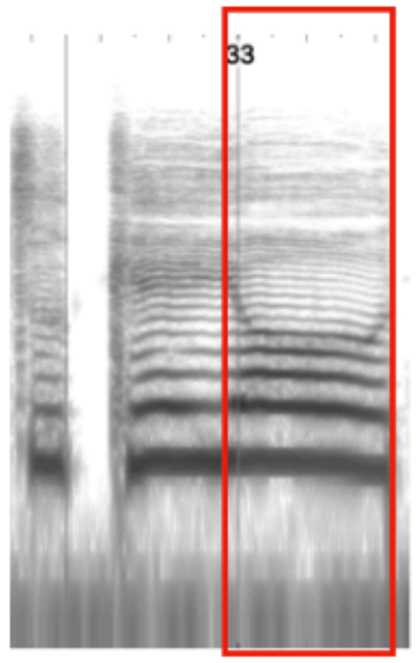
Vibrato
ビブラート



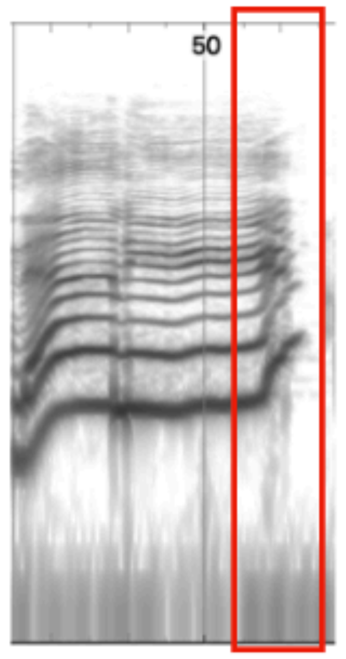
Scooping
しゃくり



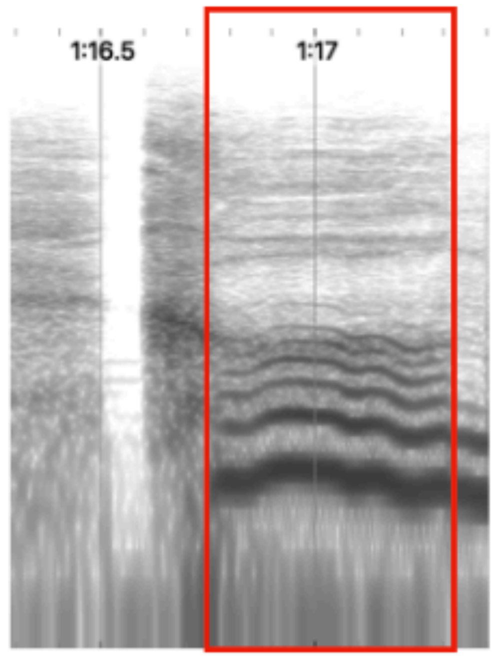
Bend
こぶし



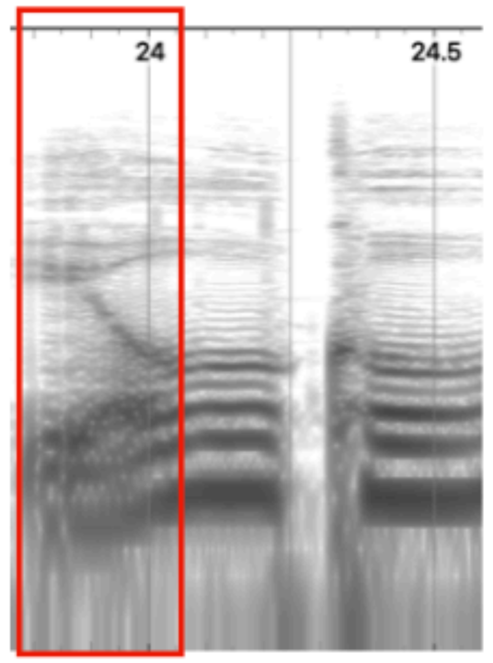
Drop
フォール



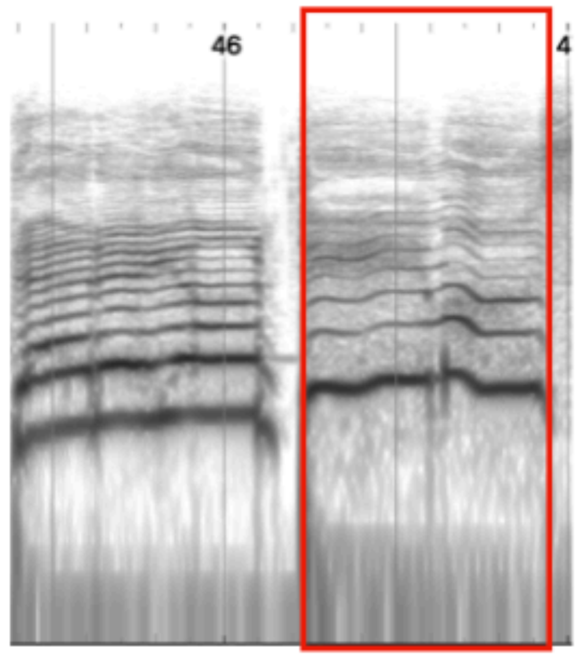
Hiccup
ヒーカップ



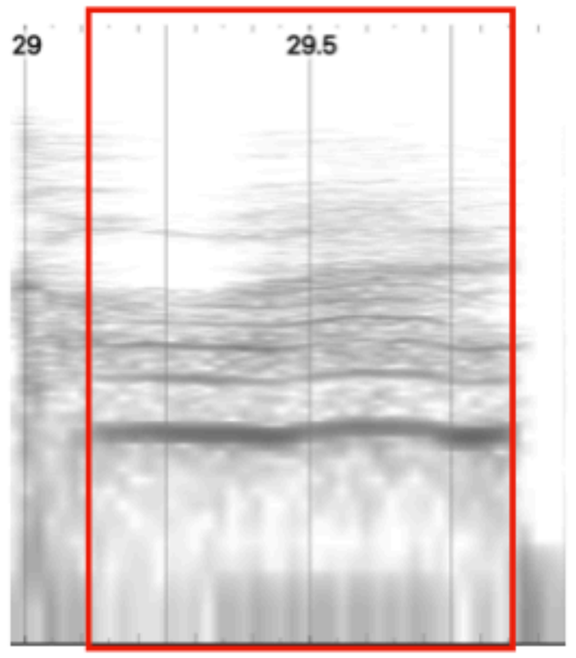
Melisma
メリスマ



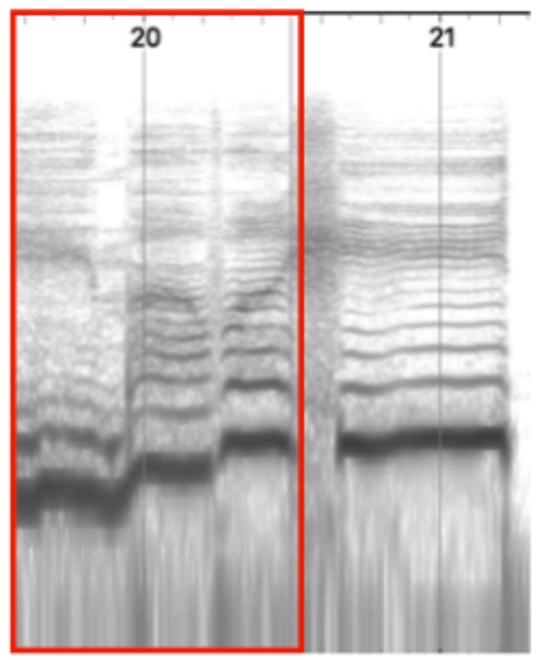
Vocal fry
ボーカル
フライ



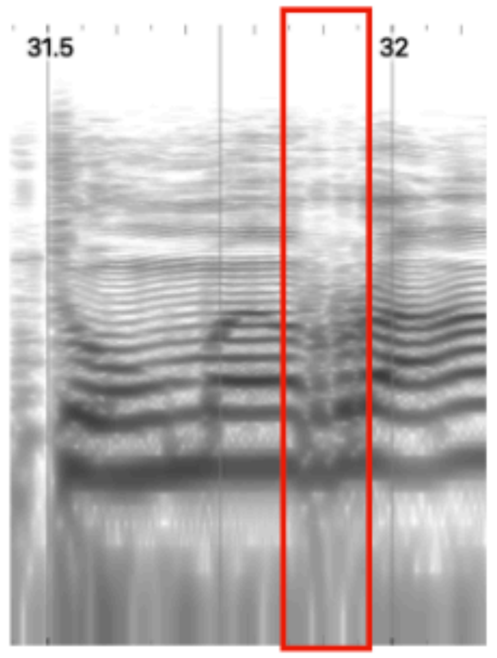
Falsetto
ファル
セット



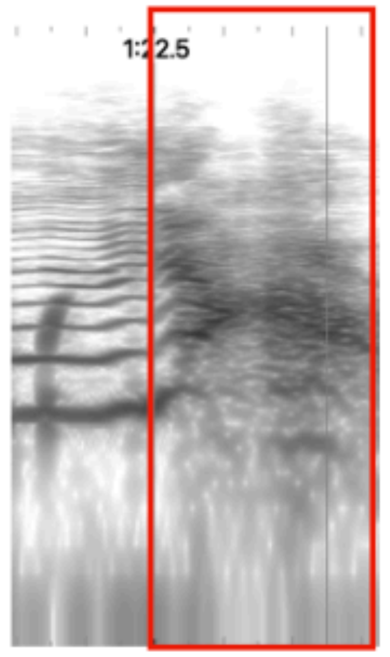
Whisper
ウィスパー



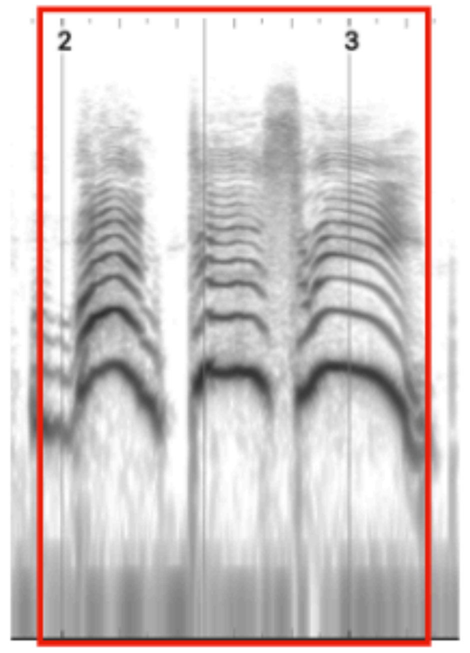
Breathy
ブレシー



Tongue trill
タング
トリル



Shout
シャウト

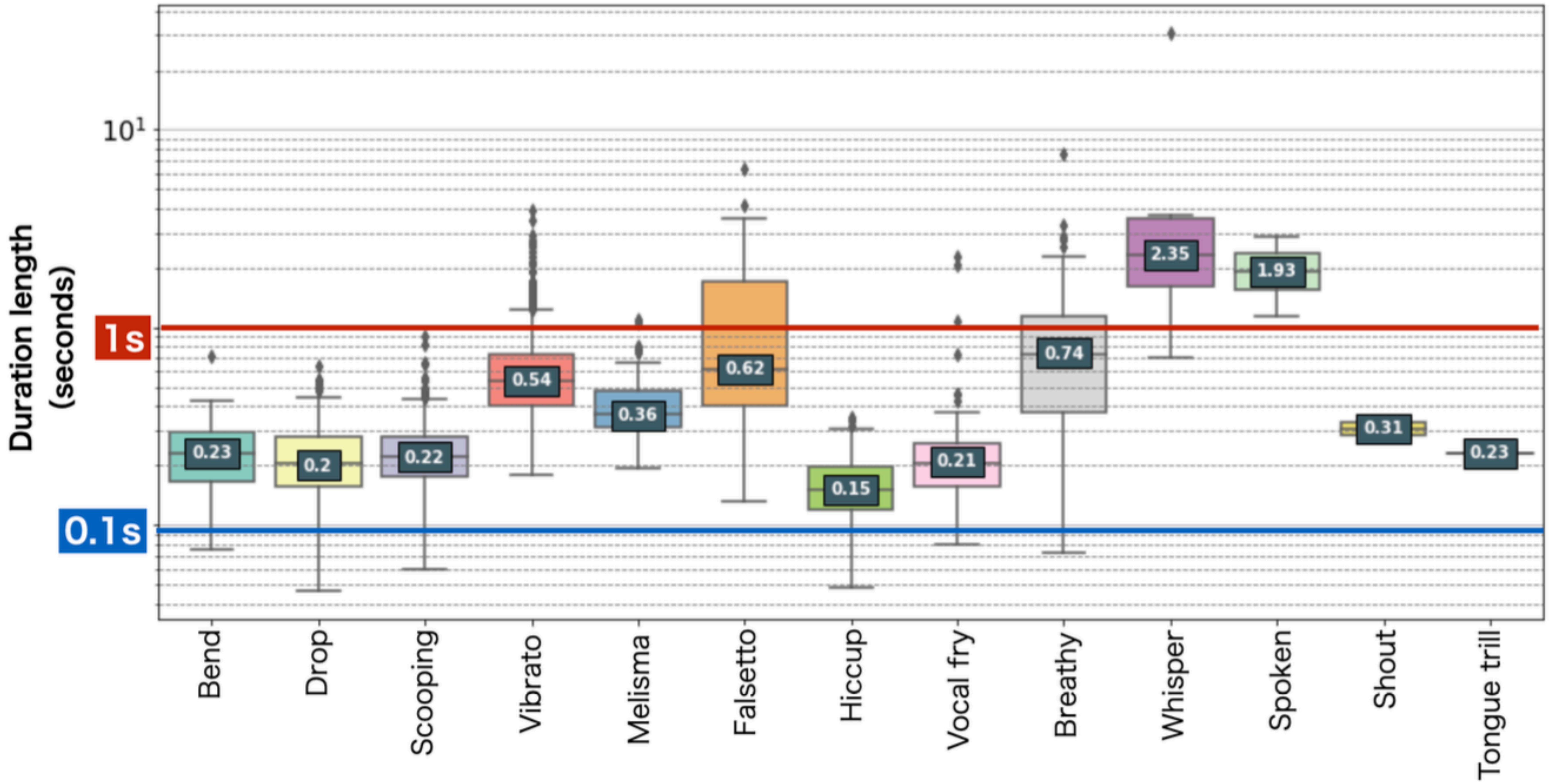


Spoken
セリフ語り

Analysis 1: Occurrence of techniques

Whole statistics

Techniques	Number of labels	Total duration [s]	Average duration [s]
Vibrato	717	448.57	0.63
Scooping	528	118.40	0.24
Bend	144	33.14	0.23
Drop	140	31.25	0.23
Hiccup	126	20.35	0.16
Melisma	38	16.36	0.44
Whisper	11	54.5	4.95
Falsetto	86	96.16	1.14
Breathy	52	41.57	1.03
Vocal fry	82	21.73	0.28
Tongue trill	1	0.36	0.23
Shout	2	1.16	0.39
Spoken	4	13.71	1.98



- Pitch techniques (vibrato, scooping, bend, drop) are frequent
- Most of techniques are short length (0.1s - 1s)

Analysis 1: Occurrence of techniques

	aiko 01	aiko 02	Takaki 01	Tamaki 02	Ayaka 01	Ayaka 02	chara 01	chara 02	Fukuyama 01	Fukuyama 02	GACKT 01	GACKT 02	Hamasaki 01	Hamasaki 02	Hirahara 01	Hirahara 02	Hirai 01	Hirai 02	Hitoto 01	Hitoto 02	YUKI 01	YUKI 02	Koda 01	Koda 02
Bend	5	5	7	0	10	1	1	0	3	3	1	0	0	0	1	1	10	3	6	4	0	0	8	6
Drop	11	11	0	0	0	0	2	1	0	1	12	12	2	0	0	0	1	2	3	3	2	1	2	3
Scooping	12	12	6	0	22	17	4	4	10	10	12	8	7	14	12	12	44	28	16	10	0	0	3	6
Vibrato	11	15	20	9	13	16	1	1	12	21	14	18	7	5	22	10	38	21	8	17	1	0	15	16
Melisma	2	0	1	0	0	0	2	1	0	3	0	0	0	0	0	0	3	2	0	0	0	0	0	0
Falsetto	3	3	1	1	4	4	0	3	0	0	2	2	1	1	0	1	2	2	0	0	0	0	2	4
Hiccup	3	3	0	1	0	1	0	0	0	0	2	1	8	2	0	0	3	1	1	0	0	0	4	0
Vocal fry	0	0	2	0	1	1	3	2	3	2	2	0	2	2	1	0	5	3	1	0	0	0	3	2
Breathy	0	0	0	1	0	1	1	1	0	1	0	0	0	0	4	3	0	1	0	4	0	0	0	0
Whisper	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0
Spoken	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Shout	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
Tongue trill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F01	F06	M03	M04	F05	F02	F02	F05	M04	M03	M01	M05	F06	F01	F03	F04	M01	M05	F05	F02	F02	F05	F06	F01

	Koyanagi 01	Koyanagi 02	hyde 01	hyde 02	Makihara 01	Makihara 02	Matsuura 01	Matsuura 02	Moriyama 01	Moriyama 02	Oda 01	Oda 02	Onitsuka 01	Onitsuka 02	Kuwata 01	Kuwata 02	Chiba 01	Chiba 02	Nishikawa 01	Nishikawa 02	Utada 01	Utada 02	Yamazaki 01	Yamazaki 02
Bend	5	5	0	2	3	11	0	2	3	2	1	2	0	0	8	4	0	1	3	9	4	1	2	1
Drop	2	3	13	11	2	2	1	1	1	1	1	6	0	1	7	3	1	0	4	5	0	2	2	2
Scooping	17	31	10	9	11	14	5	0	15	15	10	4	22	18	5	14	0	7	2	4	4	16	6	20
Vibrato	27	30	26	26	15	15	3	2	18	23	3	11	16	20	17	18	16	27	10	16	12	24	17	15
Melisma	0	0	0	0	0	1	0	0	0	0	0	0	3	1	2	0	0	0	0	0	3	5	3	6
Falsetto	0	0	4	6	0	2	3	3	8	8	0	0	0	0	0	0	0	0	0	0	7	8	1	0
Hiccup	5	0	18	9	0	0	6	17	0	0	0	0	2	0	9	0	2	9	6	9	3	0	1	0
Vocal fry	5	4	7	0	2	0	0	0	0	0	0	0	0	0	1	1	6	1	12	0	3	2	0	3
Breathy	0	4	0	1	0	1	0	0	3	0	0	8	0	1	0	1	0	0	0	0	11	3	0	2
Whisper	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Spoken	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0
Shout	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Tongue trill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	F04	F07	M03	M04	M04	M03	F01	F06	M02	M06	M02	M06	F04	F07	M06	M07	M05	M01	M05	M01	F03	F04	M06	M07

When imitating the same singer, occurrence of singing techniques is also likely to be similar？

5	11	12	11	2	3	3	0	0	0	0	0	0

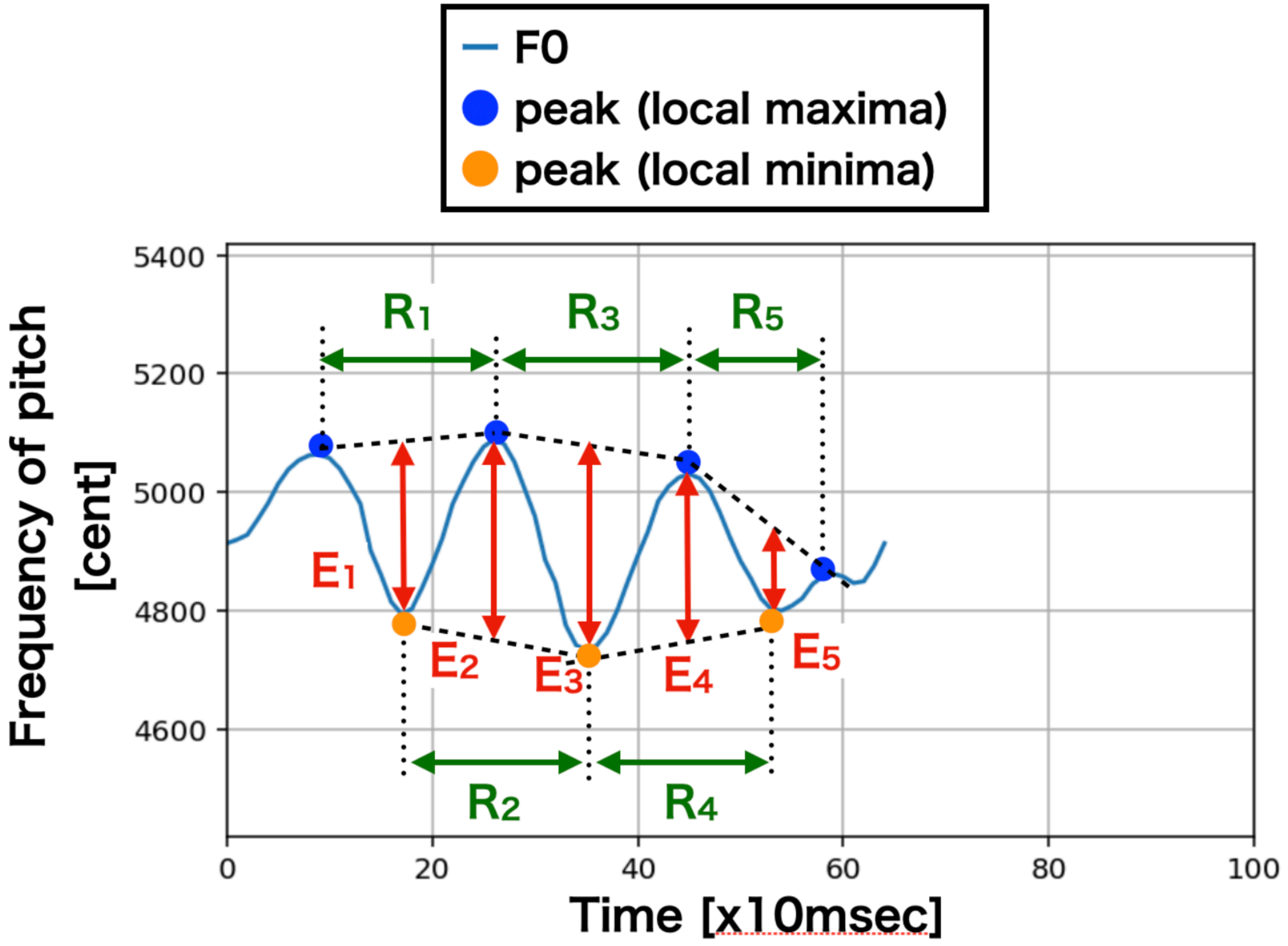
5	11	12	15	0	3	3	0	0	0	0	0	0

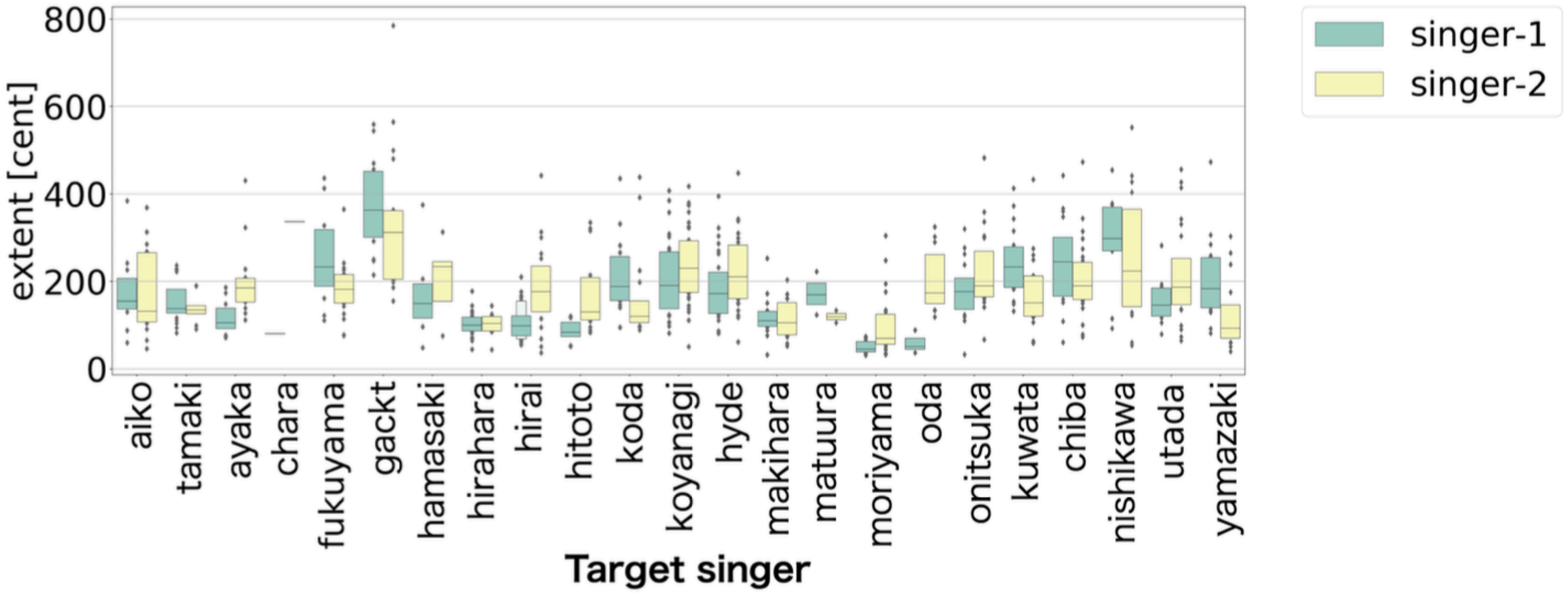
Cosine similarity
Between the occurrences
Avg. -> Intra: 0.83, Inter: 0.7

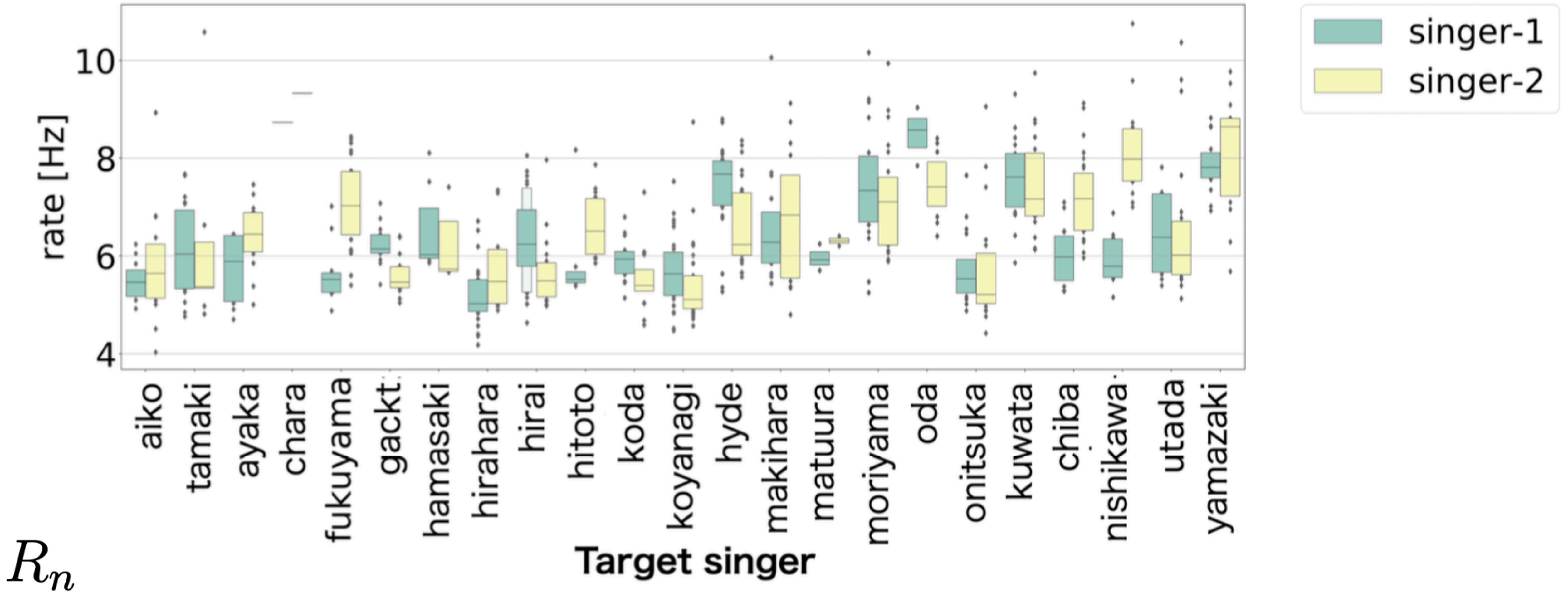
- The distribution is different by singer
- The distributions are similar between same or similar style original singer

Analysis 2. Acoustic parameters -case study of vibrato-

Picked up the vibrato labels, calculates the vibrato parameters (extent and rate)



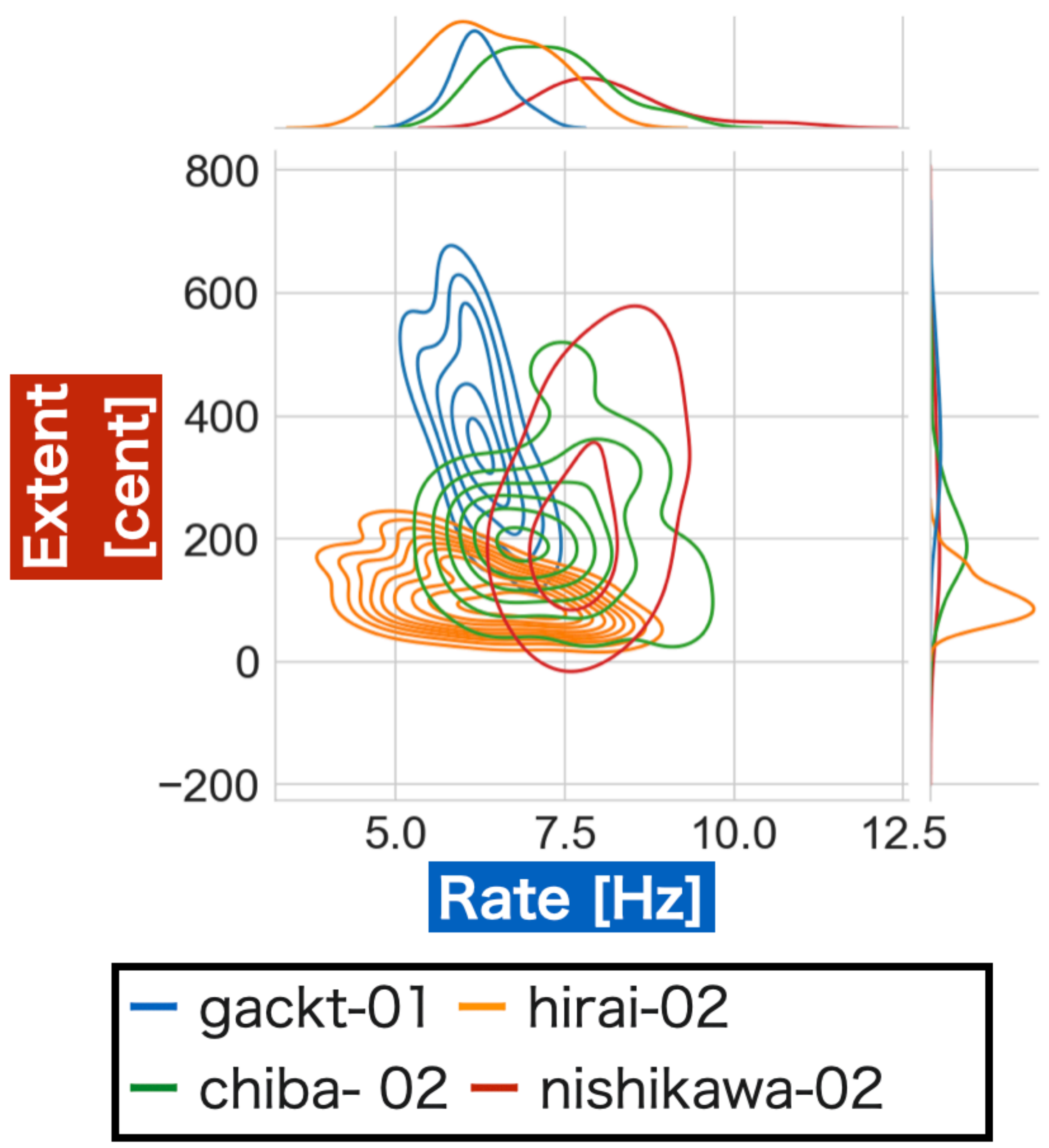


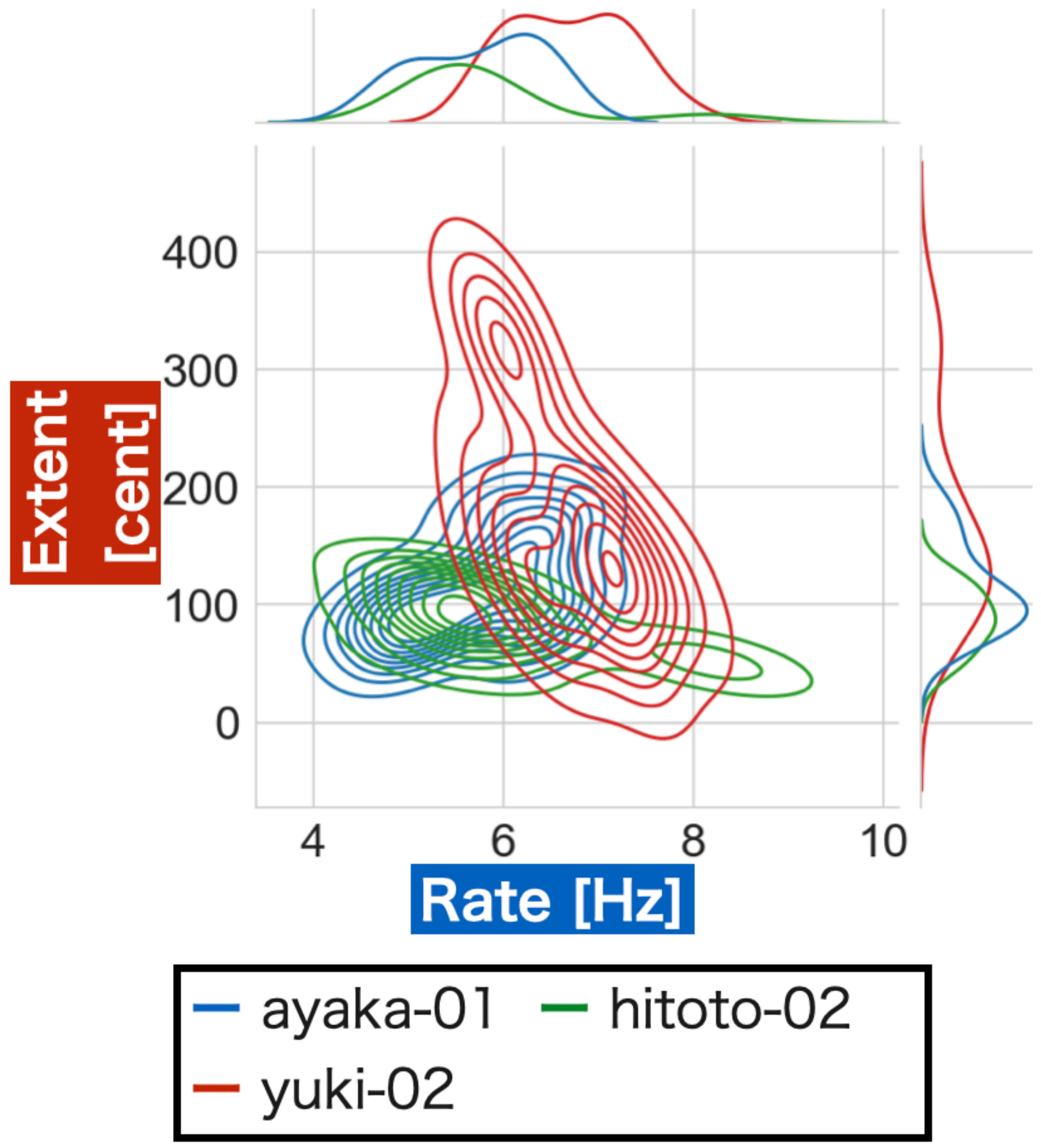


Calculation $\text{extent} = \frac{1}{N} \cdot \sum_{n=1}^{N} E_n, \quad \frac{1}{\text{rate}} = \frac{1}{N} \cdot \sum_{n=1}^{N} R_n$

Summary of discovery

Kernel distribution estimation (KDE) of rate-extent plot





The shape changes when the imitated subject is different
-> professional may be able to imitate the vibrato, in terms of the extent and rate ↔ difficult for amateur singers [Saitou 11]

Pearson's correlation coefficients with pitch and duration

	Pitch (female)	Pitch (male)	Pitch (Normalized)	Label duration
Extent	-0.424	-0.336	-0.31	-0.127
Rate	0.225	0.169	0.025	-0.268

Negative correlation between…

- Vibrato extent and pitch height (deep ∝ low)
- Vibrato rate and label duration (slow ∝ long)

· [Saitou 11] Saitou and Sakakibara. "Investigation of the Acoustic Features of Mimicry Singing and Their Impact on Perception", ASJ 2011

Analysis 3. Analysis by co-occurrence with musical elements

- **Take statistics of co-occurrence of each singing technique**
  - Associates the technique labels to each notes
  - by aggregating the label amounts

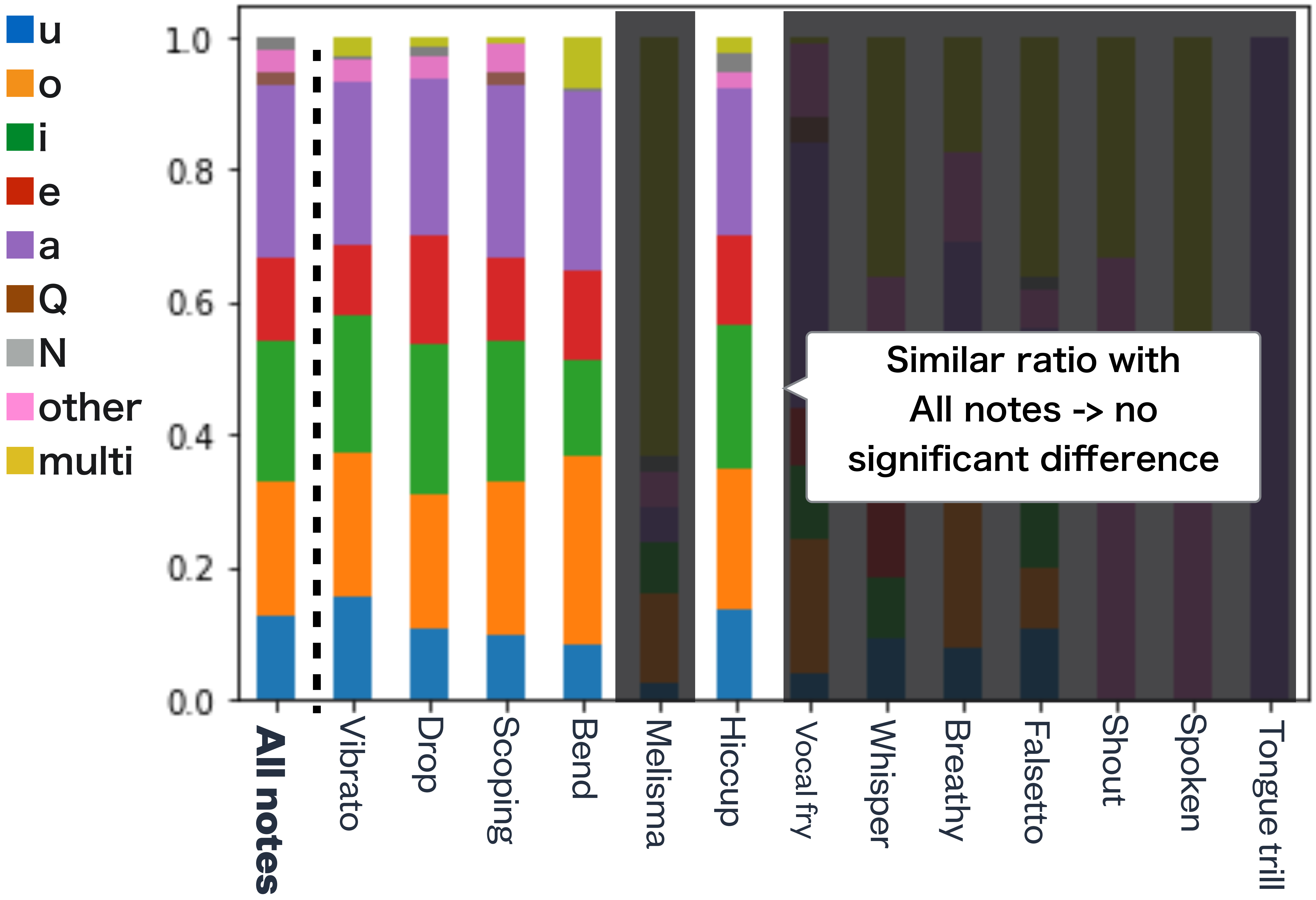
	Lyrics' vowel	Note heights	Note intervals	Duration	Phrase position
★ **Vibrato**	i x 2, u	C5, F4, Es4	+2, -1, -2	3/8, 1/8, 1/2	mid x 2, tail
★ **Bend**	o	As	As	3/8	mid

Analysis 3. Analysis by co-occurrence with musical elements

- **Lyric**：No salient correlation
- **Note heights**：More falsetto and scooping on high notes.
- **Note interval**
  - when pitch rose: more falsetto and scooping
  - when next pitch will fall: more falsetto and drop
- **Duration**
  - Many vocal fry, hiccup, bend and drop, on short notes
  - Many vibrato and scooping, on long notes.
- **Phrase position**：More vibrato on phrase end, Less Bend on phrase beginning

Vowels of note's lyrics

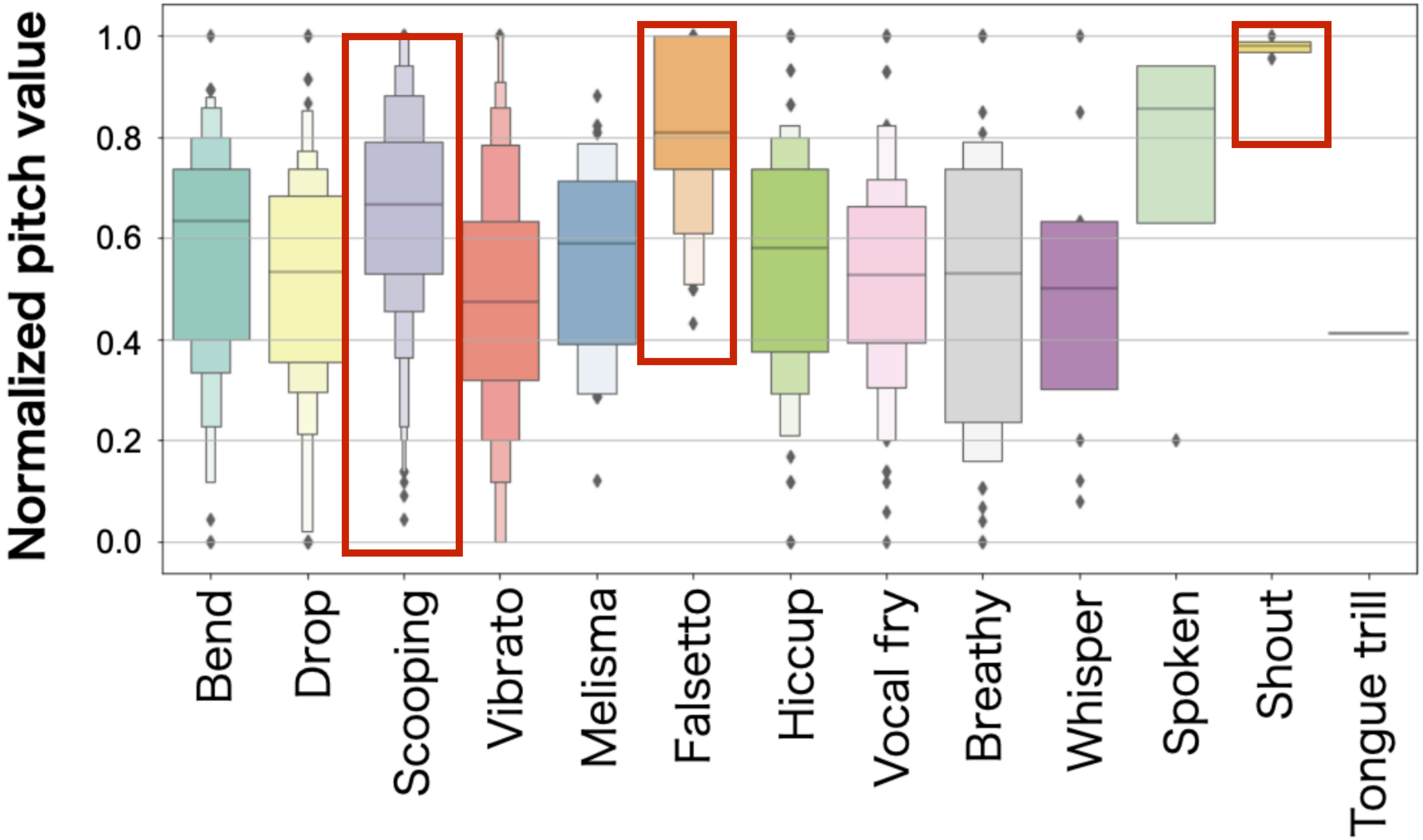
Omitted in presentation due to the time limit



Similar ratio with All notes -> no significant difference

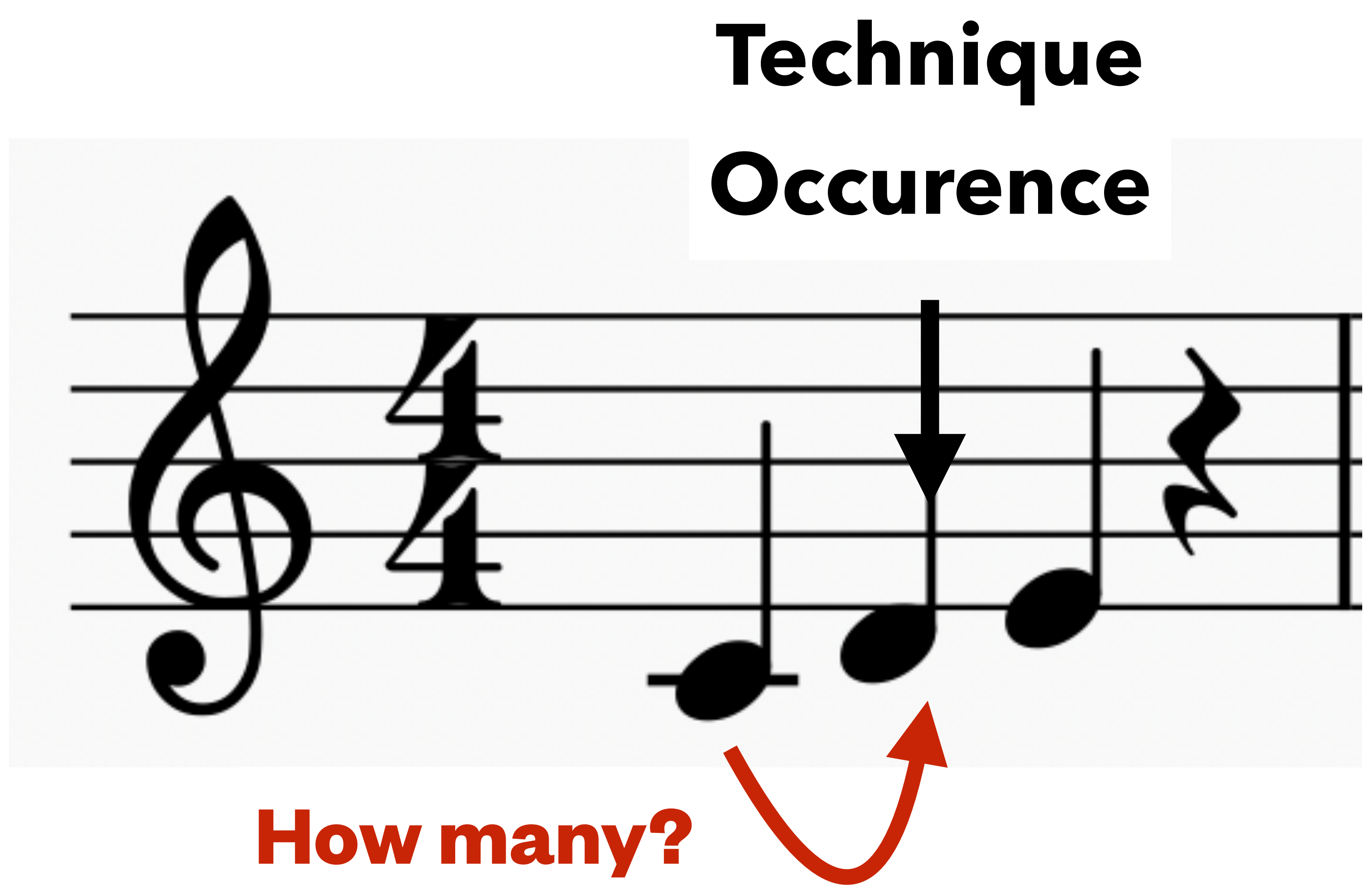
Note heights (after adjustment)

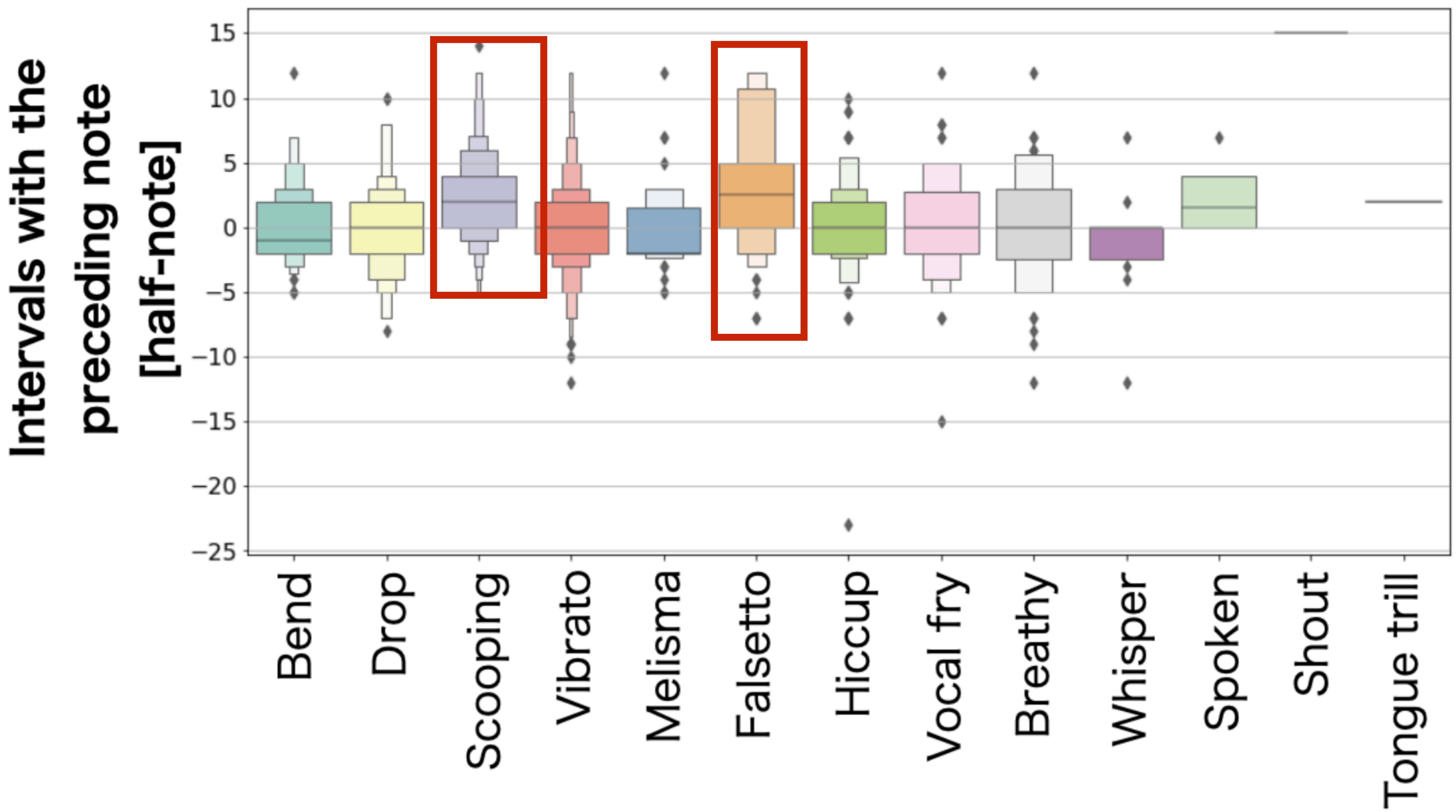
Omitted in presentation due to the time limit



Preceding note interval

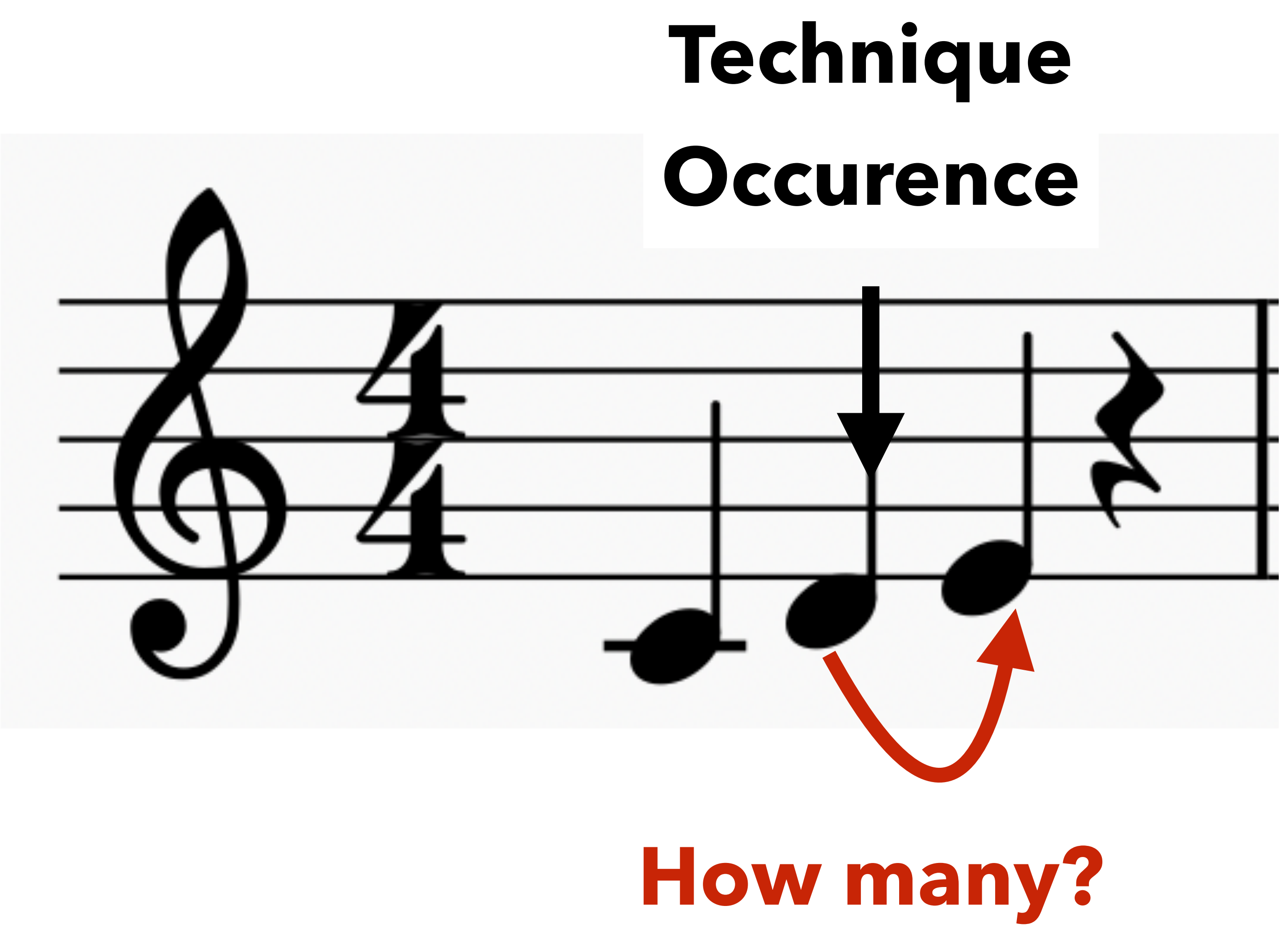
Omitted in presentation due to the time limit

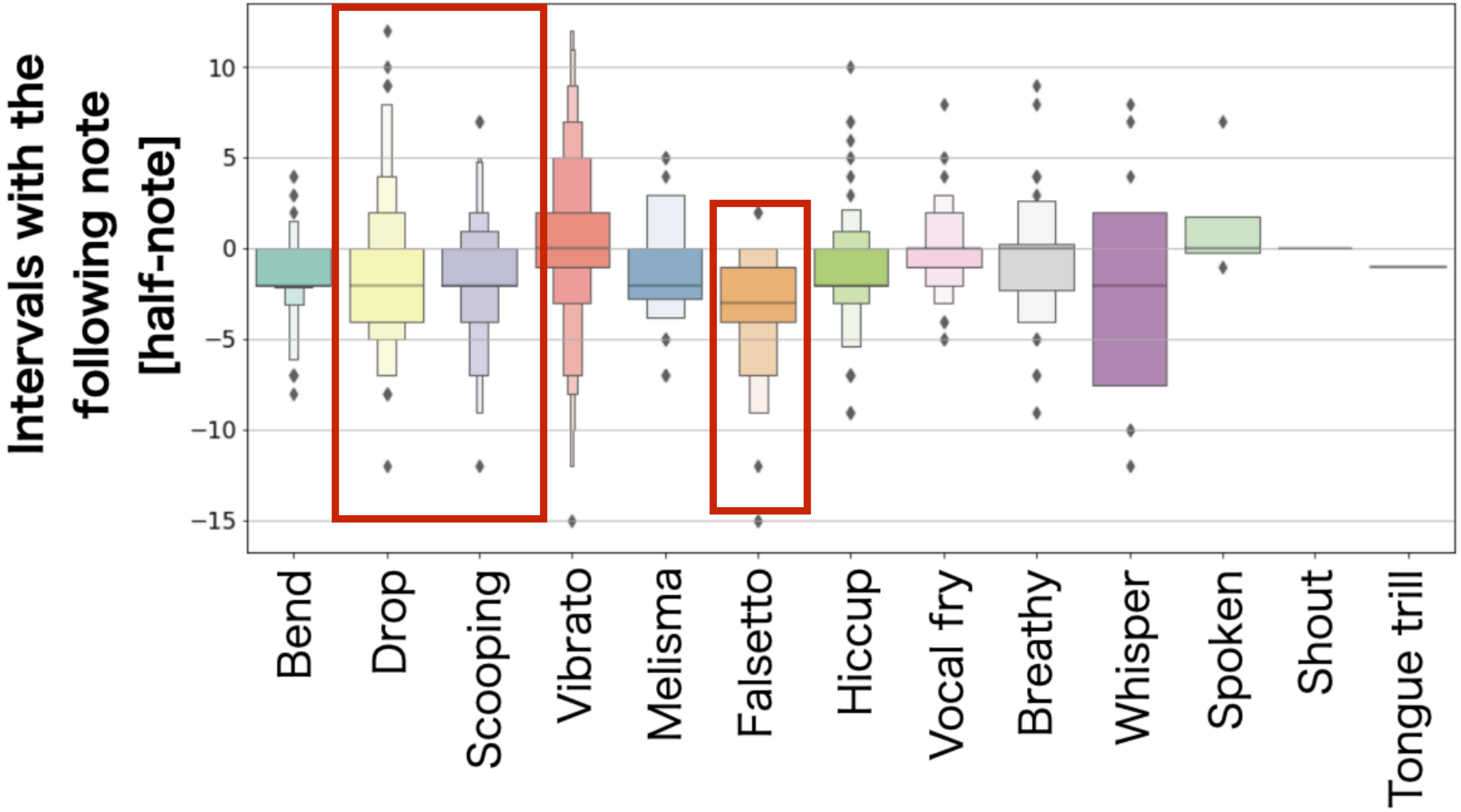




Following note interval

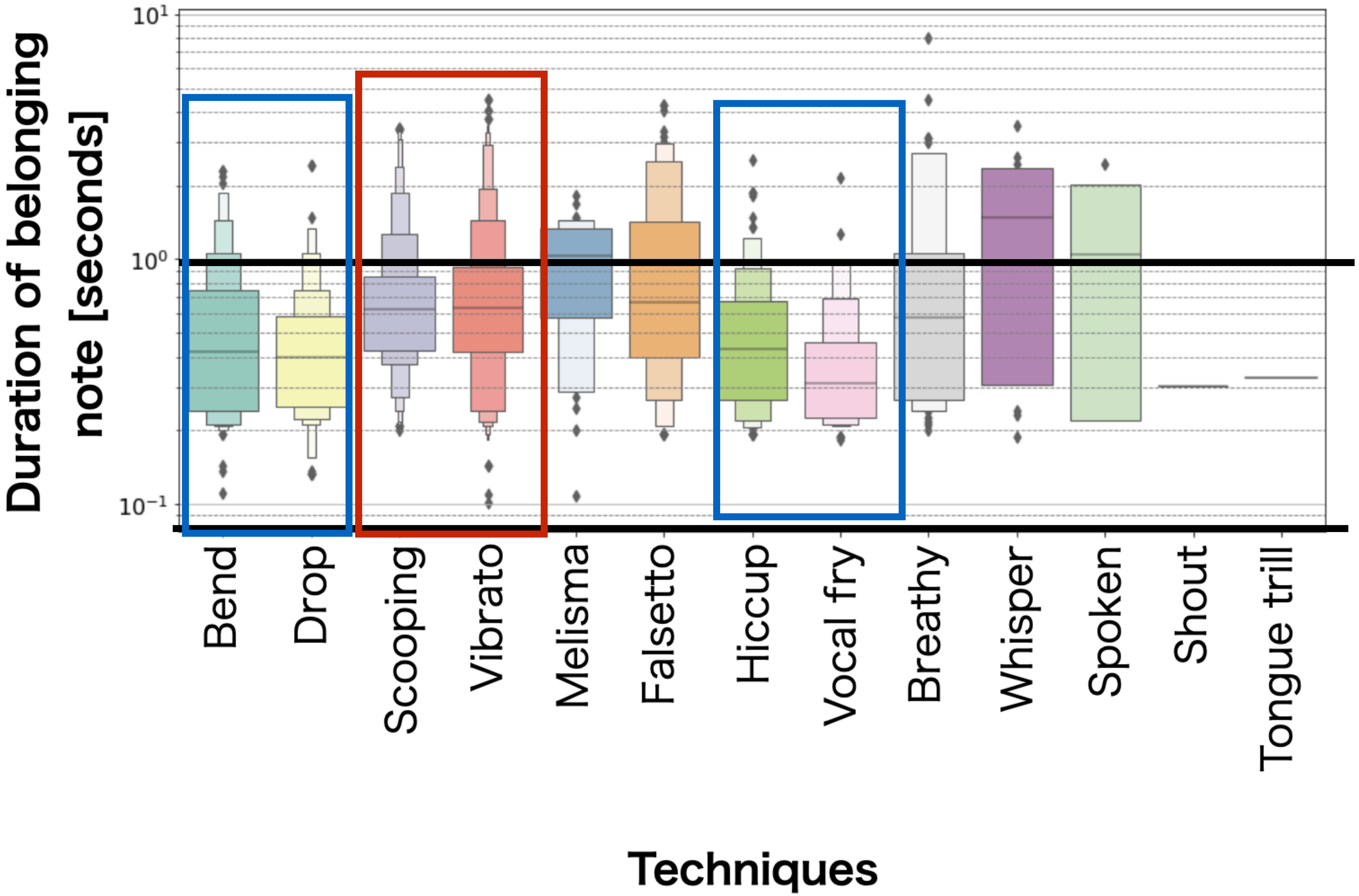
Omitted in presentation due to the time limit





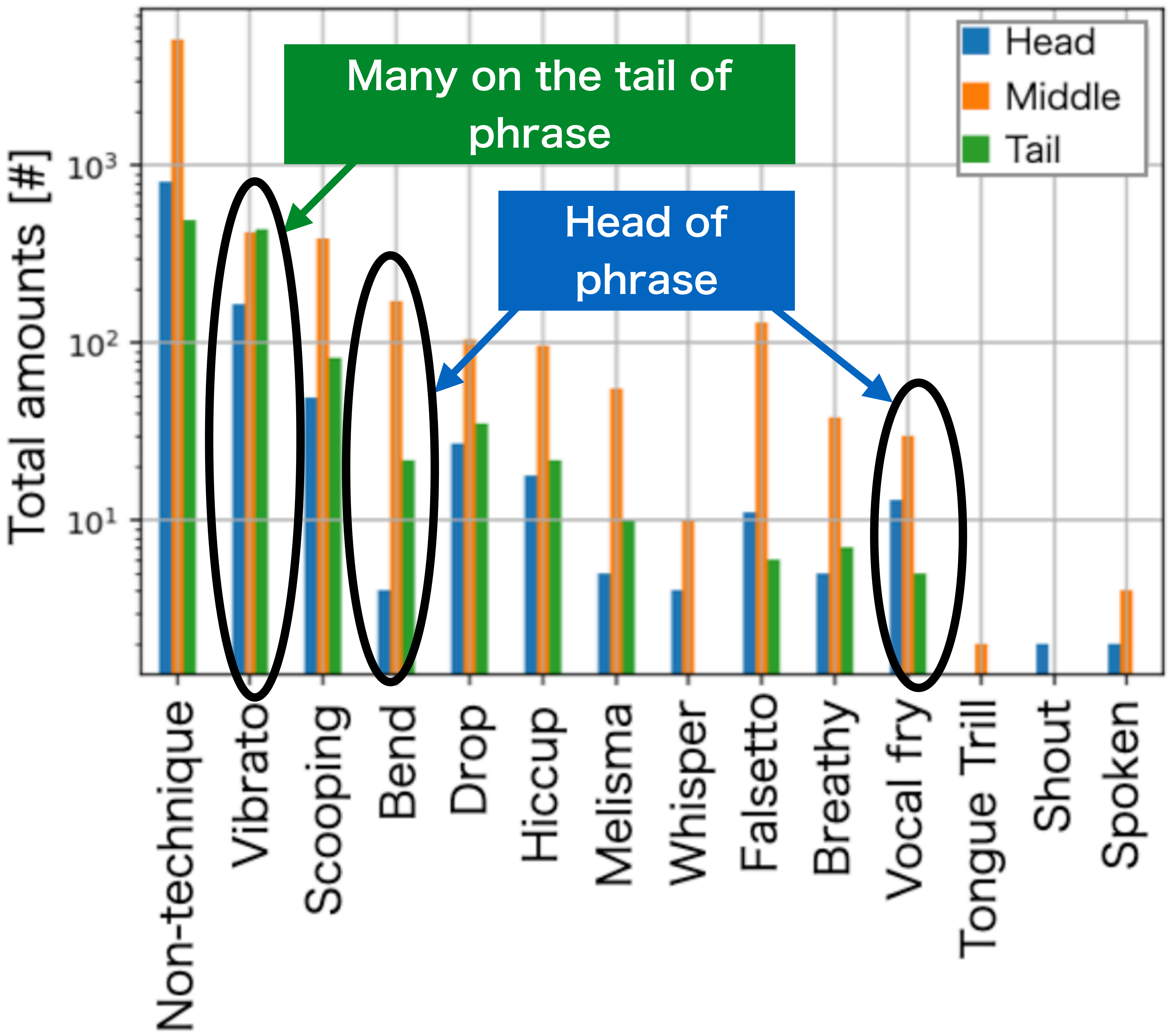
Duration

Omitted in presentation due to the time limit



Phrase position

Omitted in presentation due to the time limit



There are certain relationship between musical context and techniques

- **Occurrence：What and how often appeared?**
  - Whole analysis
  - Track-wise analysis

・Indicated occurrence and duration distribution of each technique
・Each singer has different distribution
・The distributions are similar between same or similar style original singer

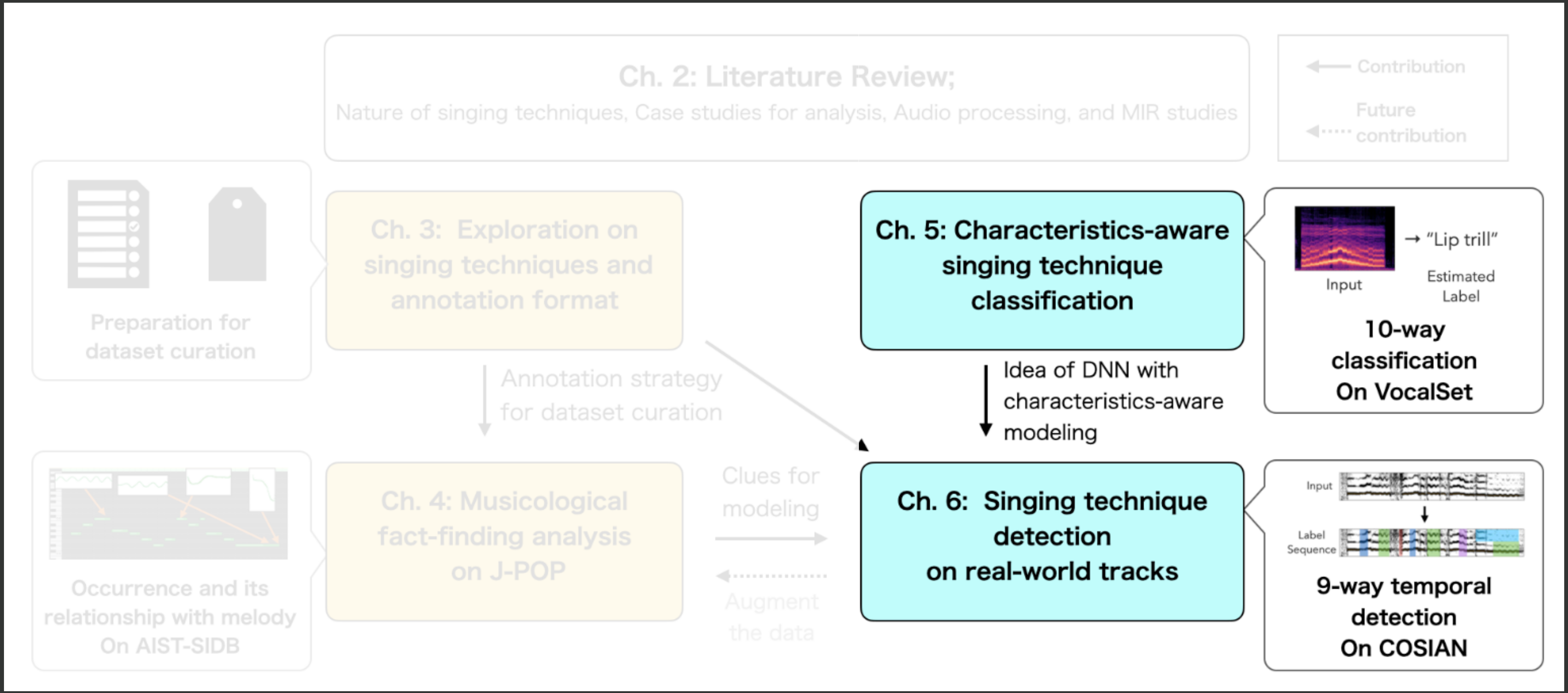
- **Vibrato parameter：How to realize vibrato?**
  - Analysis of vibrato parameters (depth and speed)
  - Differences of vibrato parameters in same imitator

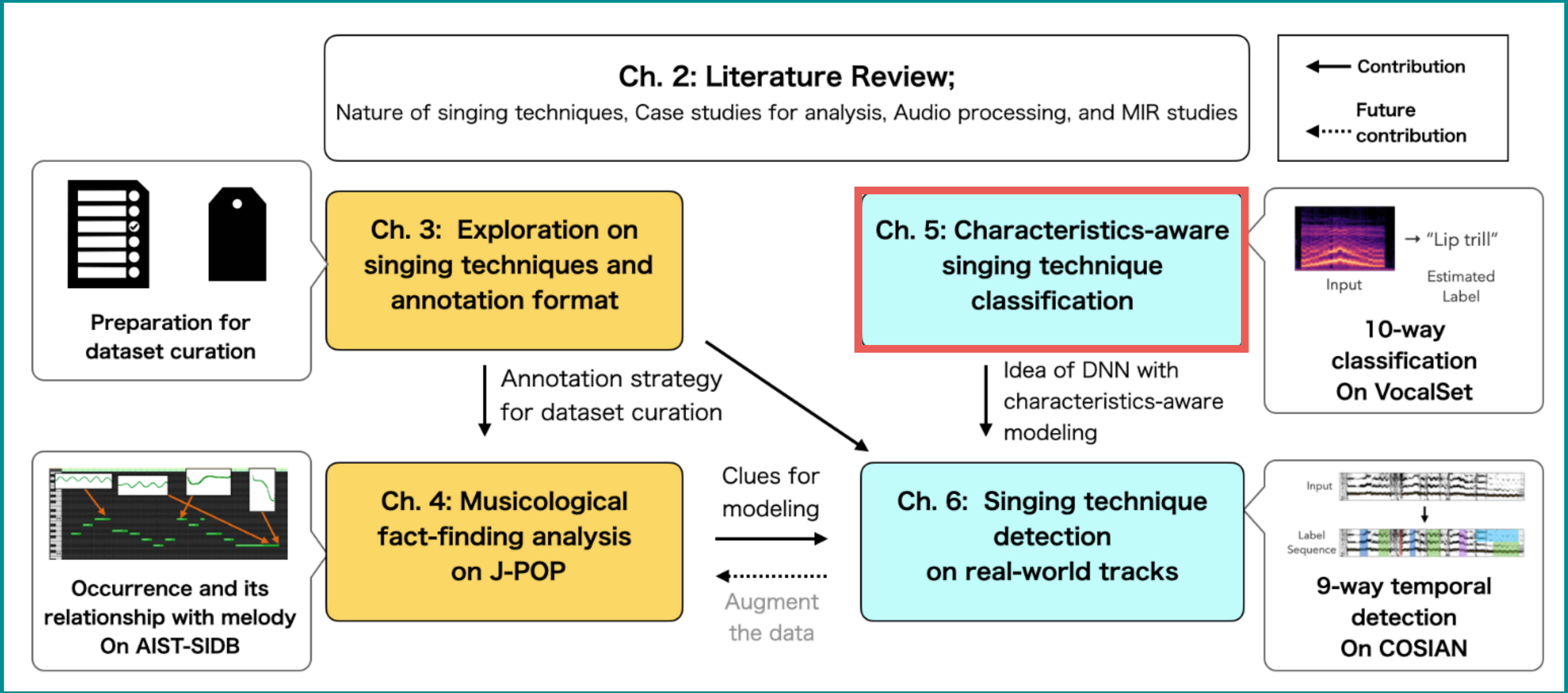
・The average depth is 181.1 cent (≒ 2 semitones)
・The average speed is 6.53 Hz
・Observed the phenomena that the singers modify the parameters based on target style

- **Location：When techniques occur?**
  - Correlation between lyrics, previous and next pitch, note duration and position in the phrase
  - Correlation between occurrence location and vibrato parameters

・lyric：No salient correlation
・pitch：More falsetto and scooping on high notes.
・pitch interval：when pitch rises: more falsetto and scooping, when next pitch will fall: more falsetto and drop
・duration：**Many vocal fry, hiccup, bend and drop, on short notes.**
Many vibrato and scooping, on long notes.
・position：More vibrato on phrase end, Less Bend on phrase beginning
・parameters：The higher the note, the shallower the vibrato, and the longer the vibrato, the slower it tends to be

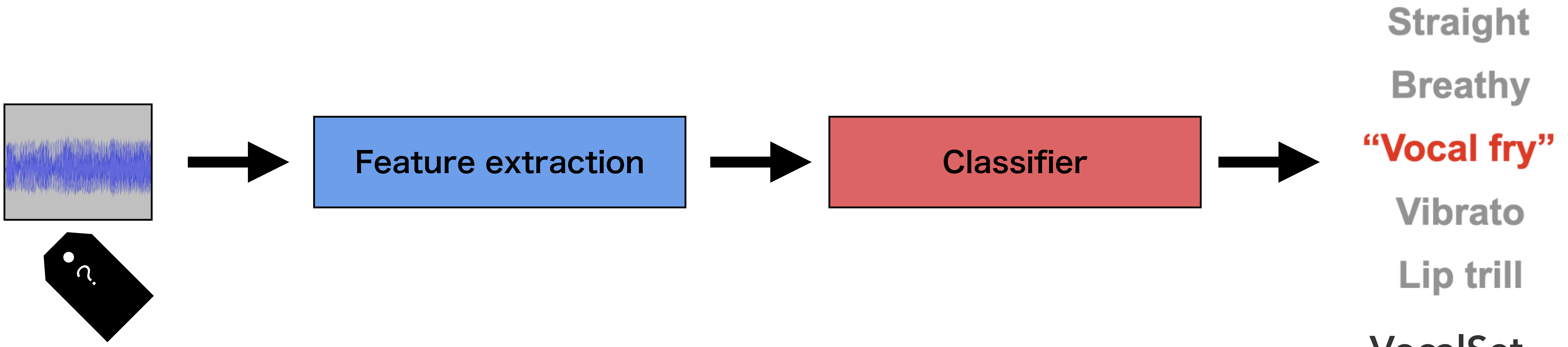
Part II. Computation





Chapter 5
Singing Technique Classification considering Feature Extraction and Imbalance-aware Learning

Exploration of singing technique classification methods



VocalSet
10-way cls.

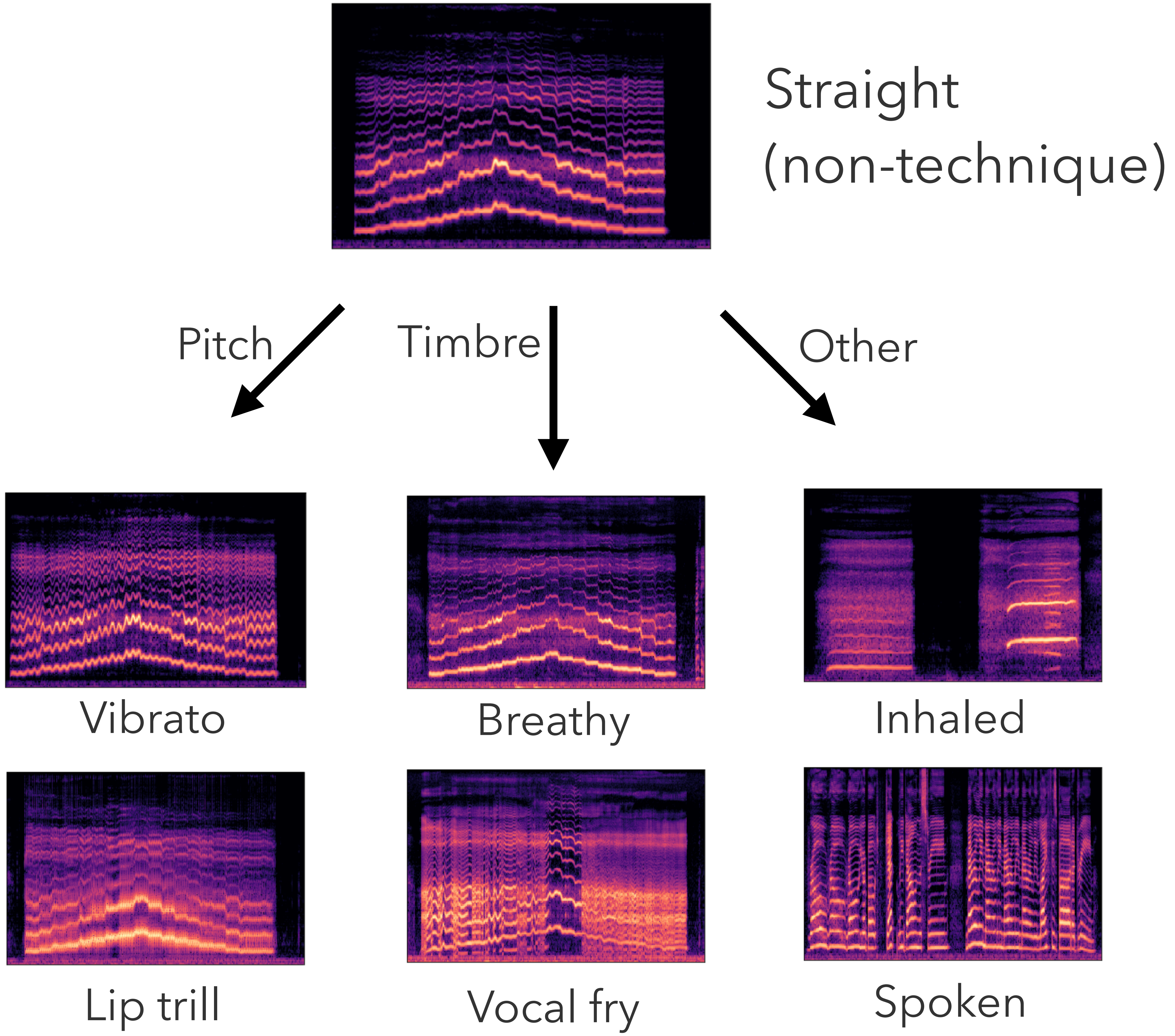
- Challenges:

  1. Feature extraction -> Which approach is better to model singing techniques?
  2. Imbalanced data -> How to mitigate bad effect from label imbalance?

Task: singing technique classification

- **Single label classification [Wilkins 18]**
  - An emerging task in music classification
  - Given audio, identify which singing techniques
  - Dataset: VocalSet

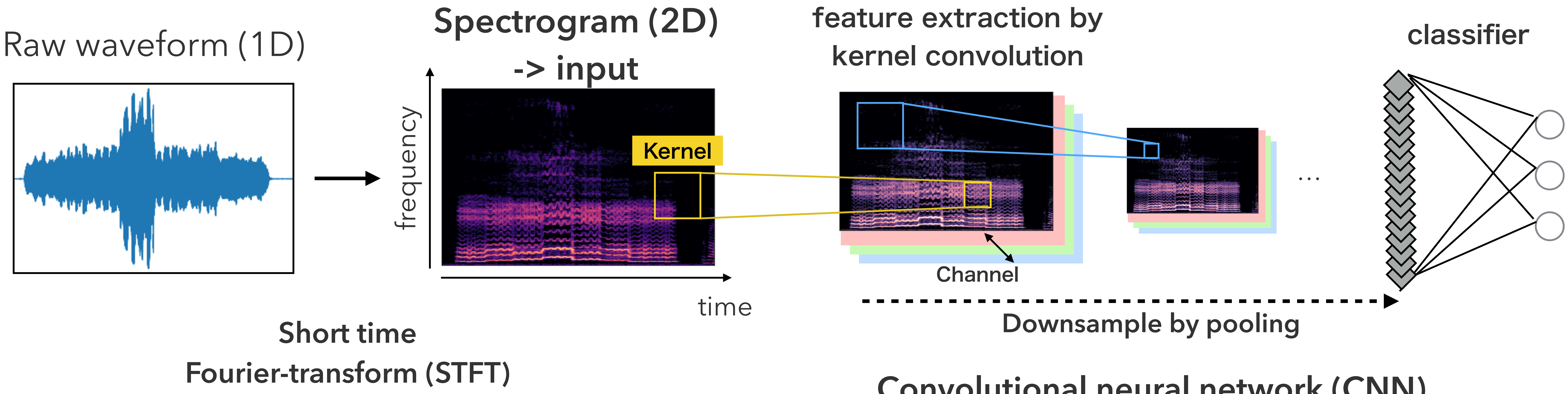
- **Challenges**
  - **1. Many fluctuation components**
    - Pitch - vibrato, lip trill etc.
    - Timbre - breathy, vocal fry etc.
    - Other - inhaled, spoken etc.
  - **2. Imbalance data**
    - Some techniques are few, as real-world



[Wilkins 18] J. Wilkins et al. “VocalSet: A Singing Voice Dataset” ISMIR 2018

Basic idea: Convolutional Neural Network (CNN), on Audio

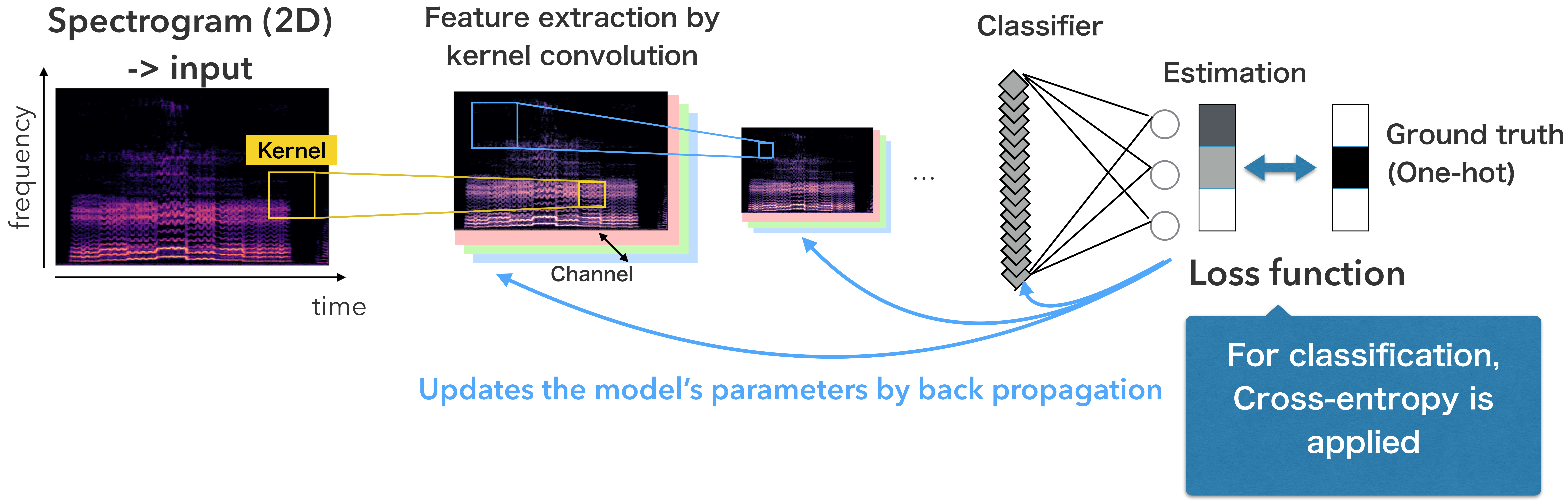
- **Making spectrogram -> process as 2D image and apply CNN**
  - Now, most common method for DNN-based audio classification [Purwins 19]



-> Expects powerful performance on feature extraction

[Purwins 19] H. Purwins et al.,¨ "Deep learning for audio signal processing," IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 2, pp. 206–219, 2019

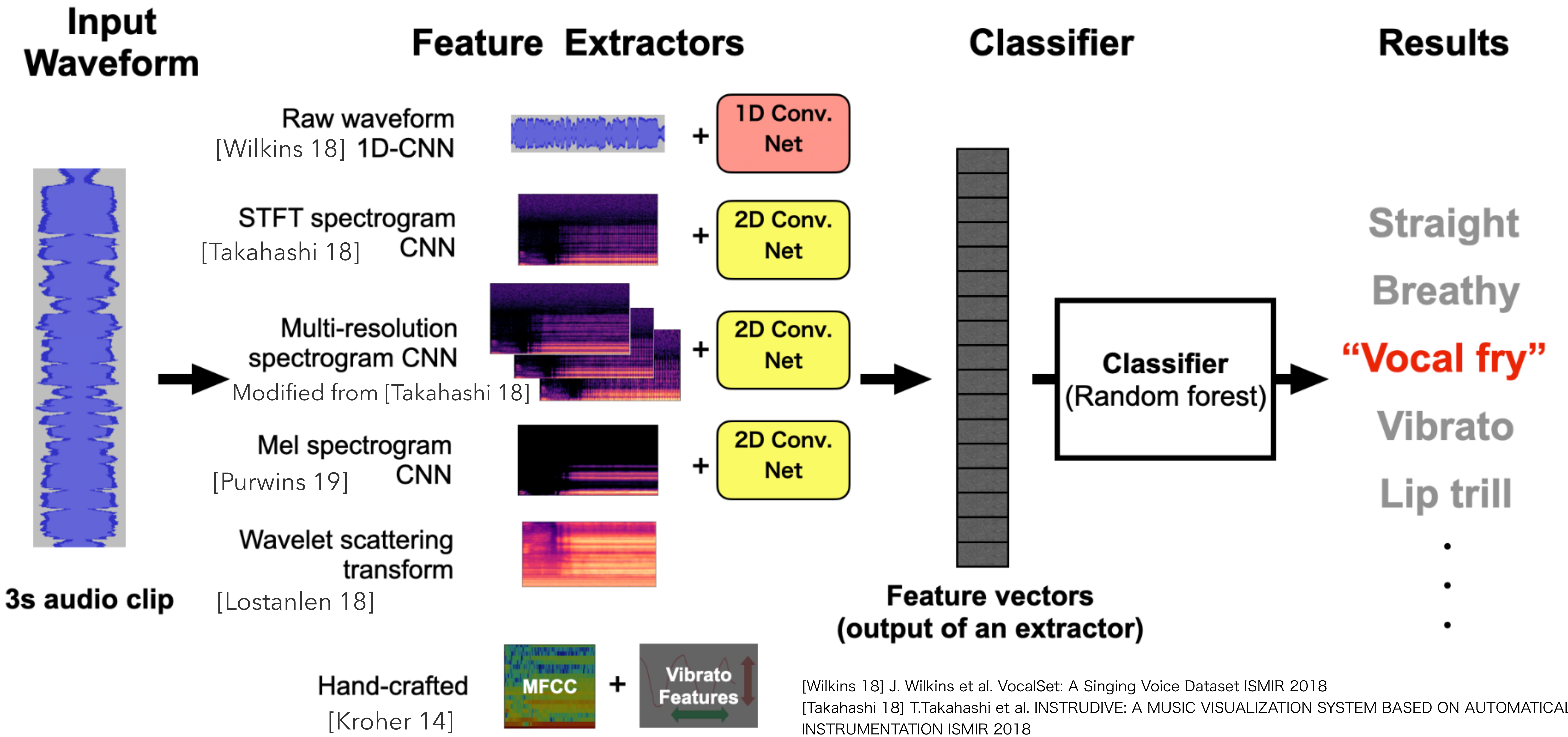
- Based on the minimization of loss function, updates the parameters



[Purwins 19] H. Purwins et al., ̈ "Deep learning for audio signal processing," IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 2, pp. 206–219, 2019

① Which audio representation is better?

Compared various setting including hand-crafted feature and CNN-based learning



[Wilkins 18] J. Wilkins et al. VocalSet: A Singing Voice Dataset ISMIR 2018

[Takahashi 18] T.Takahashi et al. INSTRUDIVE: A MUSIC VISUALIZATION SYSTEM BASED ON AUTOMATICALLY RECOGNIZED INSTRUMENTATION ISMIR 2018

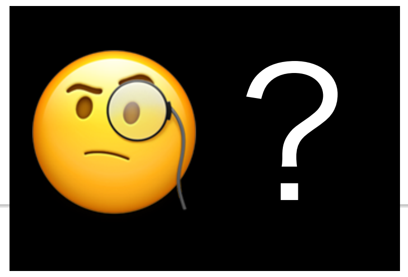
[Purwins 19] H. Purwins et al.Deep Learning for Audio Signal Processing, IEEE JSTSP 2019

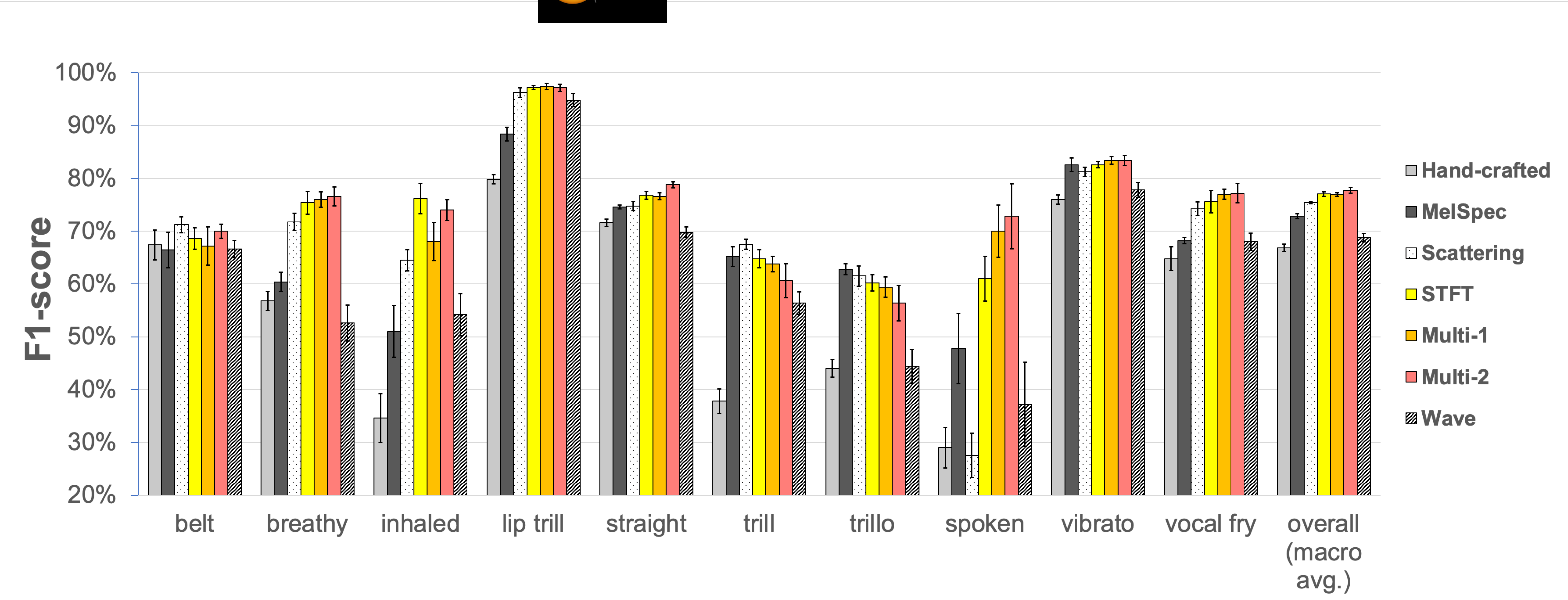
[Lostanlen 18] V. Lostanlen et al. Extended playing techniques: the next milestone in musical instrument recognition, DLfM 2018

[Kroher 14] N. Kroher et al. Improving accompanied Flamenco singing voice transcription by combining vocal detection and predominant melody extraction. ICMC/SMC 2014.

① Which feature extraction is better?

Multi-resolution spectrogram + 2D Oblong-kernel CNN achieved the best performance

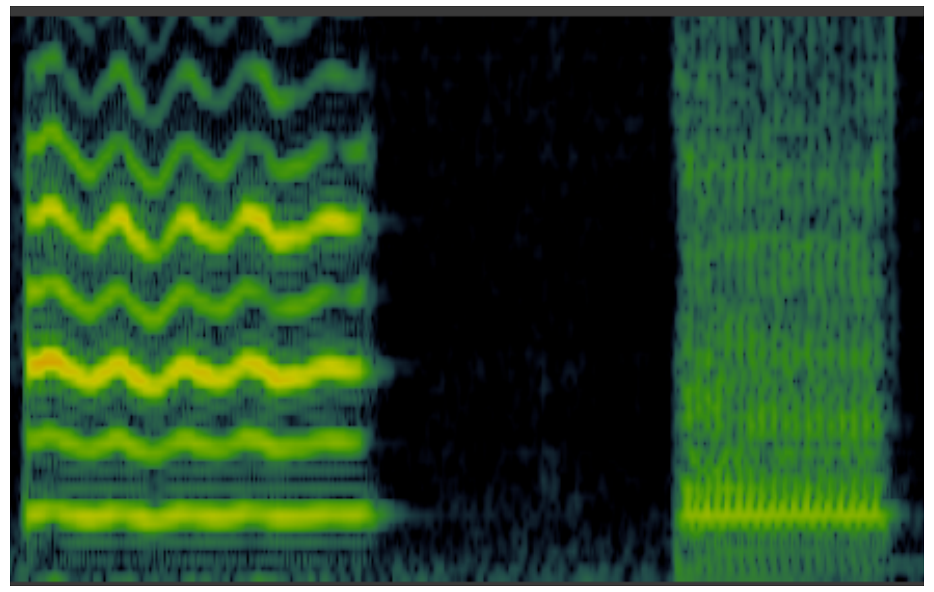




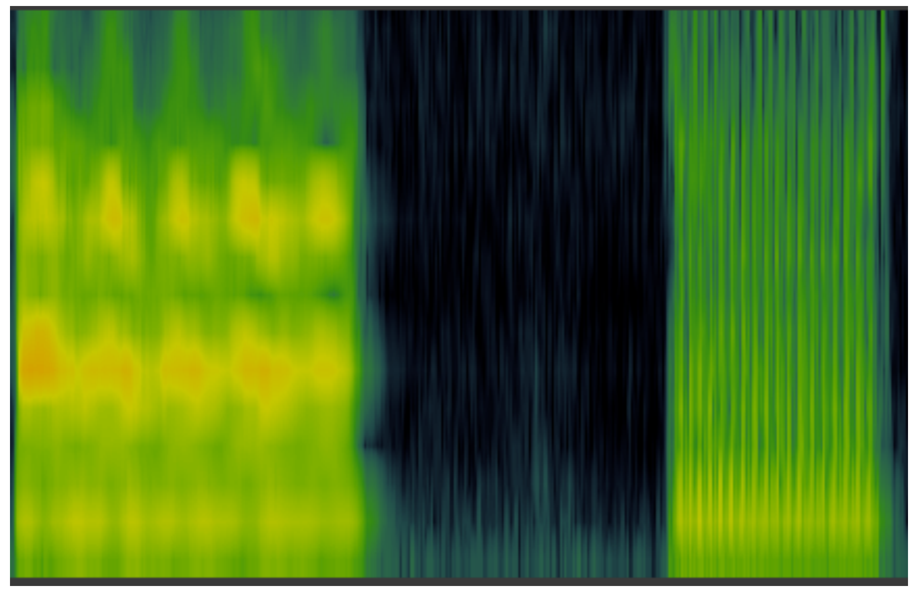
Modification for Input and CNN

- **Input representation -> Multi-resolution**
  - Stacked 3 different resolution spectrograms on channel axis, by differentiating FFT-size at STFT
  - Expects to adapt various fluctuation

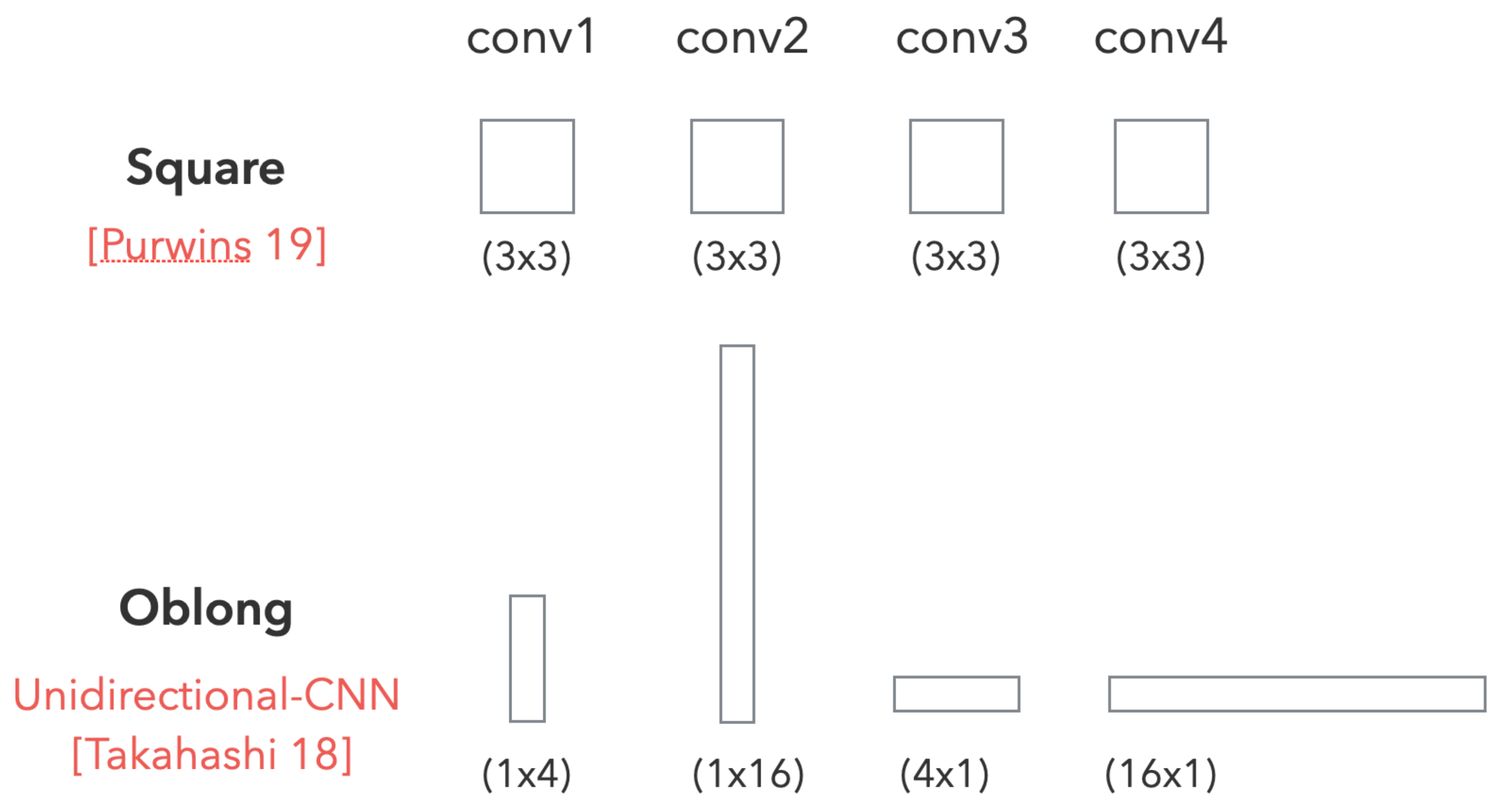
- **CNN -> Adapts the kernel shape**
  - Modified oblong-shaped kernel, from 3x3 square-shaped kernel, which is widely used in CNN
  - Expects to capture more meaningful locality-contexts
  - Vertical long -> timbral feature, Horizontal long -> temporal feature



FFT-length: long (2048)
frequency: high
time: low



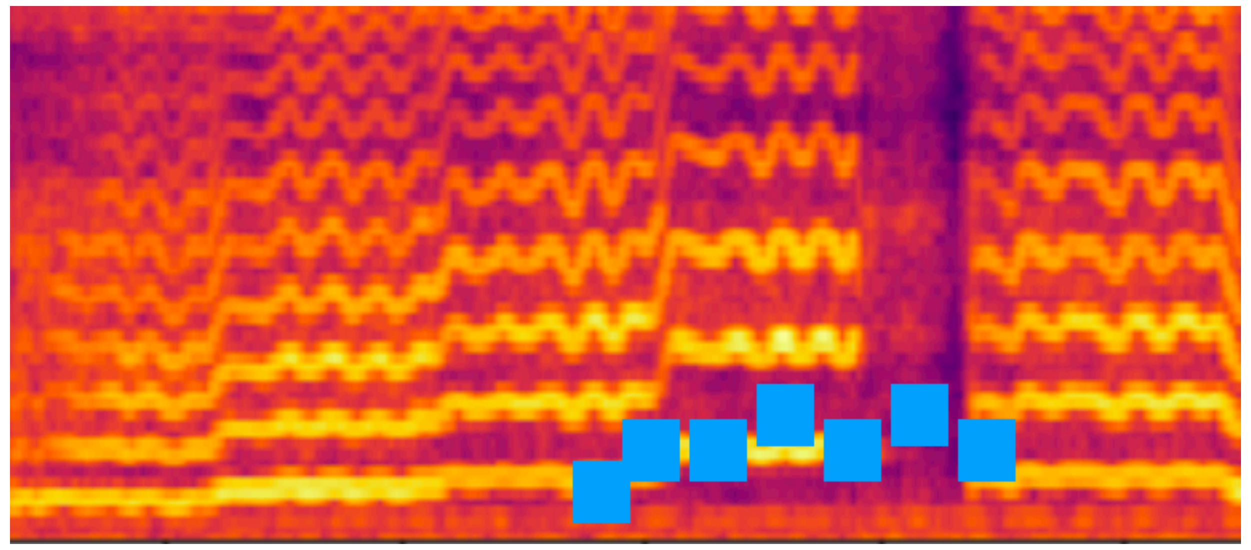
FFT-length: short (512)
frequency: low
time: high



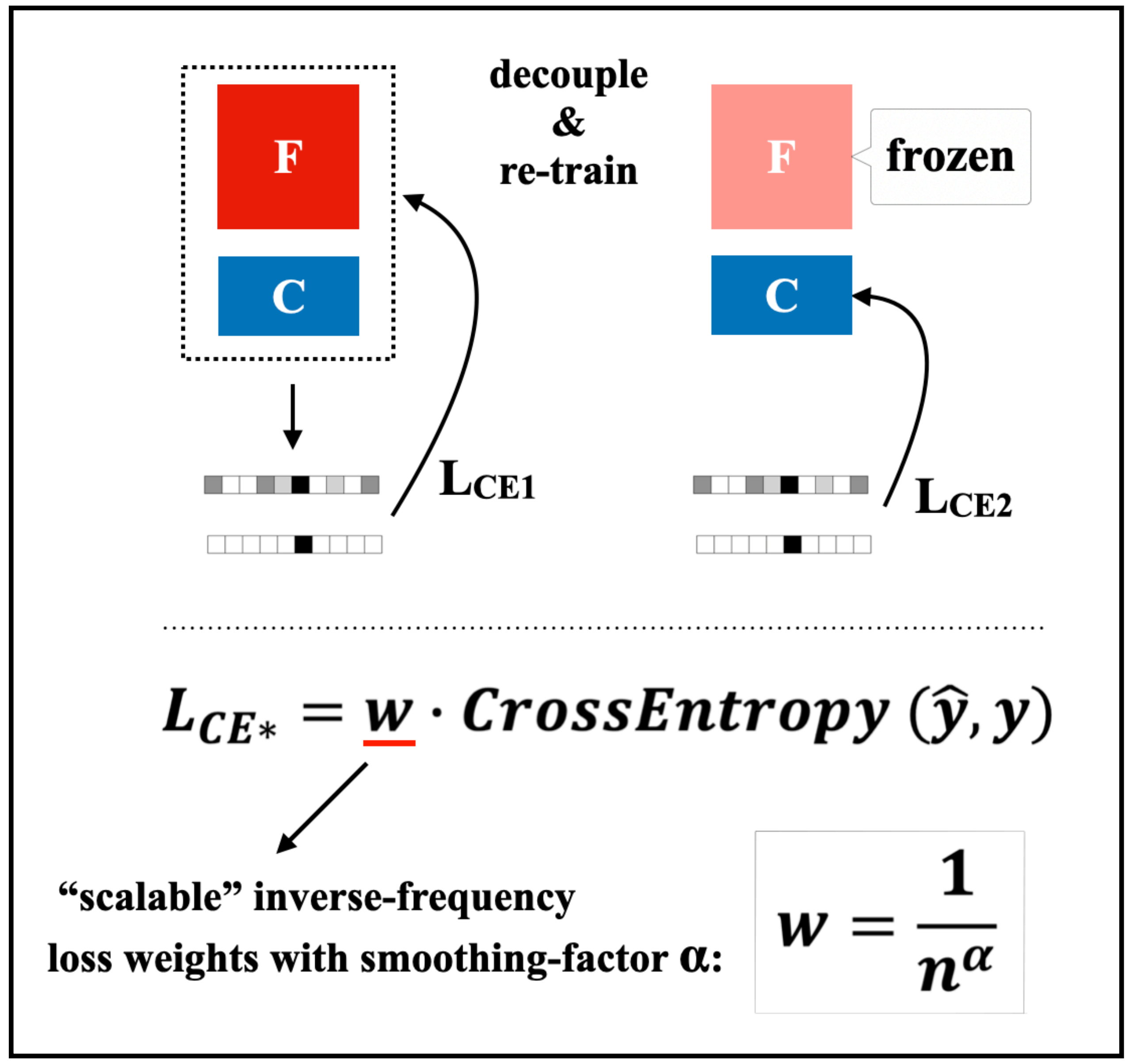
② Further characteristics-aware modeling

- **1. Adapt Deformable convolution [Dai 17]**
  - Dynamically determined the kernel shape

  →Expects to adapt more on various fluctuation
  - "Singing technique exhibits the geometric patterns on the spectrogram"



- **2. Classifier retraining (cRT) with inverse frequency weight [Kang 20]**
  - 1st: train entire part, 2nd: only retrain the classifier part
  - Adopts inverse-frequency-weighted cross entropy for loss function. (Scaling factor α=0.2)



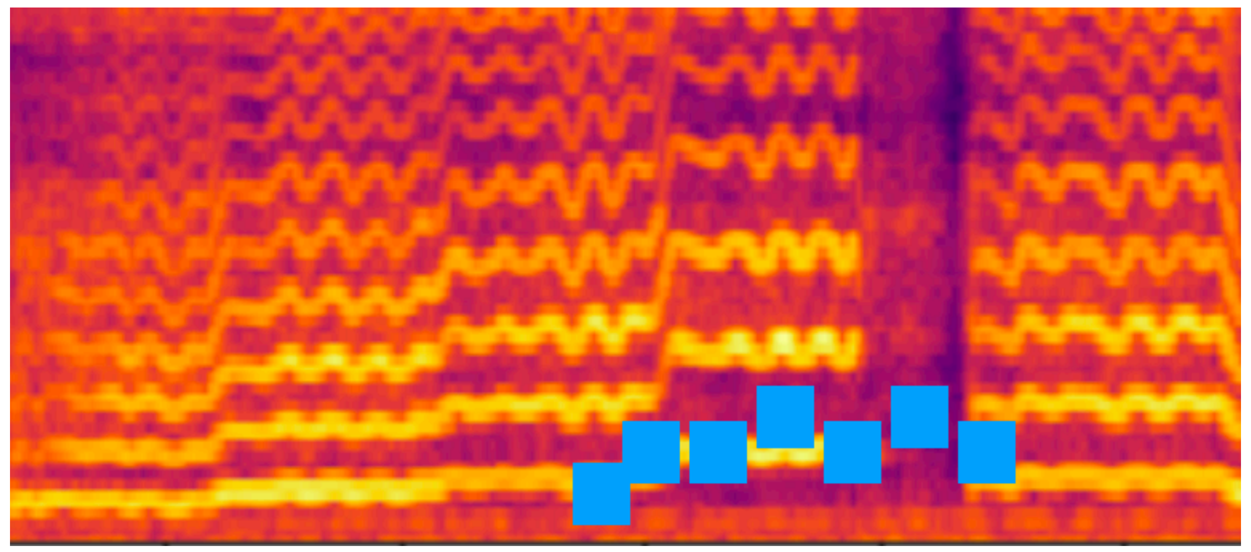
[Dai 17] Deformable Convolutional Networks J. Dai et al., ICCV 2017
[Kang 20] Decoupling Representation and Classifier for Long-Tailed Recognition. B. Kang et al, ICLR 2020

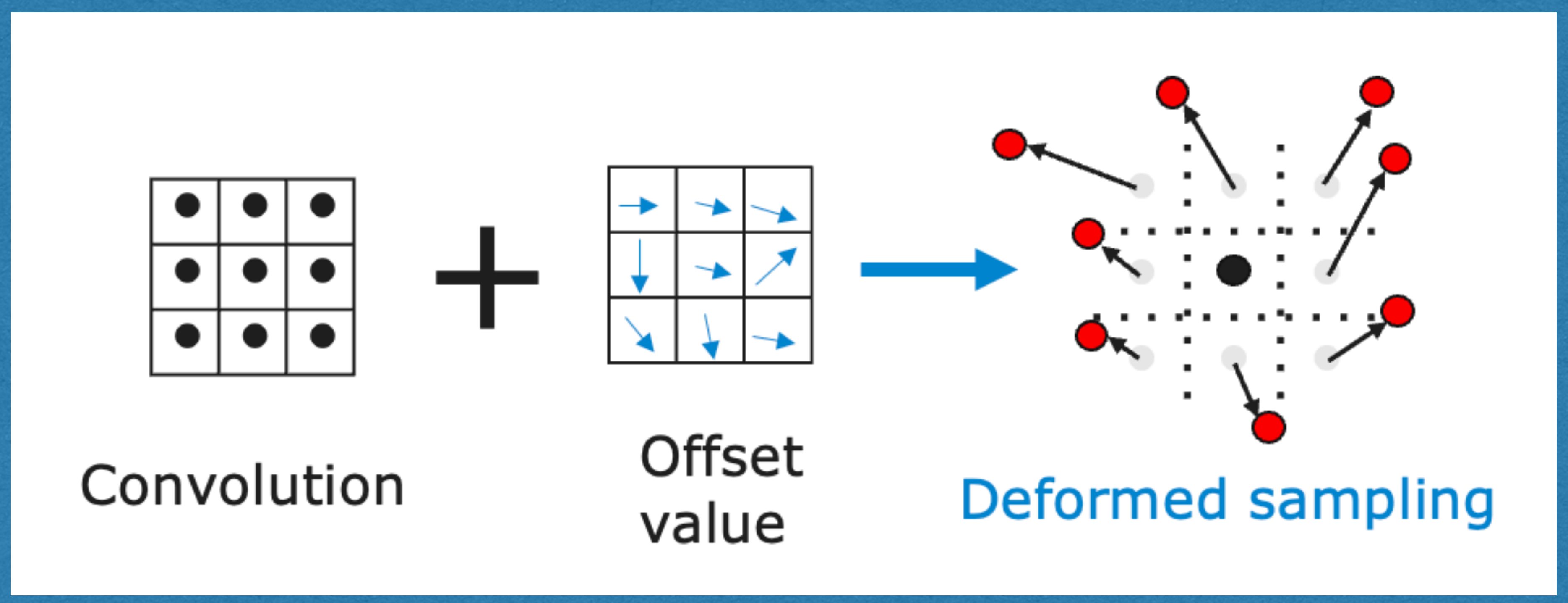
② Further characteristics-aware modeling

- **1. Adapt Deformable convolution [Dai 17]**
  - Dynamically determined the kernel shape

→Expects to adapt more on various fluctuation

  - “Singing technique exhibits the geometric patterns on the spectrogram”





- **2. Classifier retraining (cRT) with inverse frequency weight [Kang 20]**
  - 1st: train entire part, 2nd: only retrain the classifier part
  - Adopts inverse-frequency-weighted cross entropy for loss function. (Scaling factor α=0.2)

$$L_{CE*} = w \cdot CrossEntropy(\hat{y}, y)$$

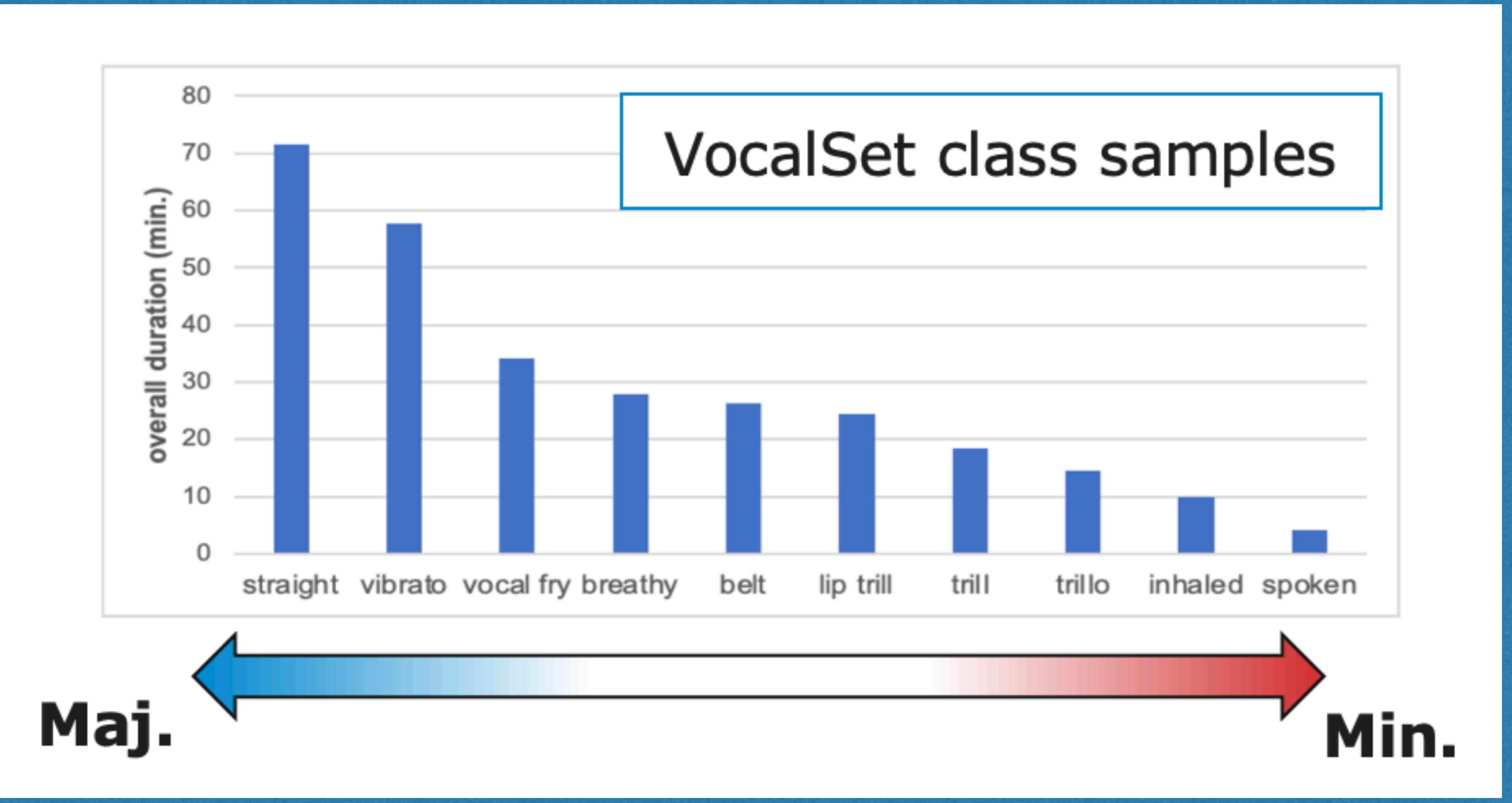
“scalable” inverse-frequency loss weights with smoothing-factor α: $w = \frac{1}{n^{\alpha}}$

[Dai 17] Deformable Convolutional Networks J. Dai et al., ICCV 2017
[Kang 20] Decoupling Representation and Classifier for Long-Tailed Recognition. B. Kang et al, ICLR 2020

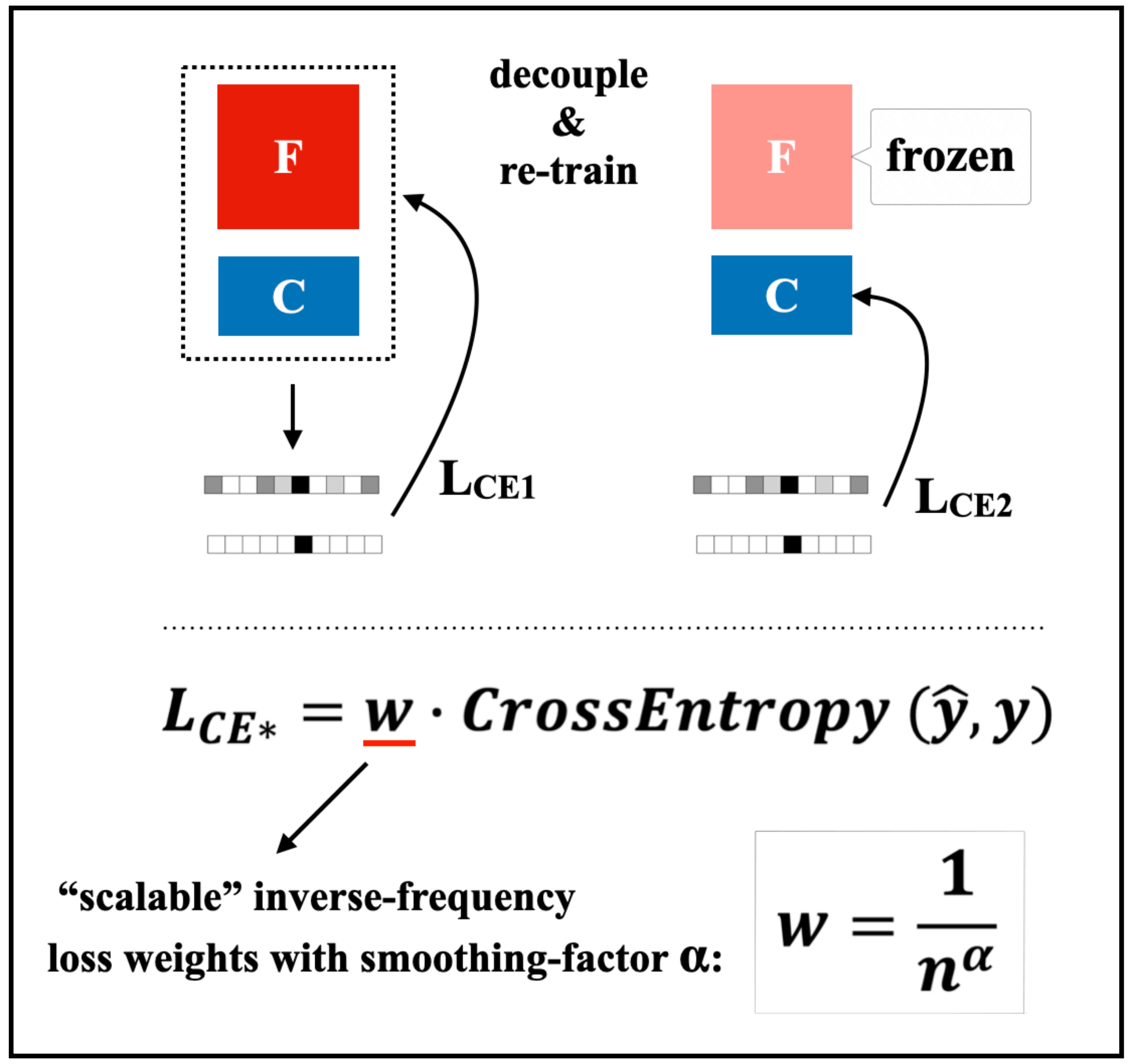
② Further characteristics-aware modeling

- 1. Adapt Deformable convolution [Dai 17]
  - Dynamically determined the kernel shape

  →Expects to adapt more on various fluctuation
  - "Singing technique exhibits the geometric patterns on the spectrogram"



- **2. Classifier retraining (cRT) with inverse frequency weight [Kang 20]**
  - 1st: train entire part, 2nd: only retrain the classifier part
  - Adopts inverse-frequency-weighted cross entropy for loss function. (Scaling factor α=0.2)



[Dai 17] Deformable Convolutional Networks J. Dai et al., ICCV 2017
[Kang 20] Decoupling Representation and Classifier for Long-Tailed Recognition. B. Kang et al, ICLR 2020

Results of the full model

※unlike the comparison part, the split is by singer, Since the official split is found after the comparison

+ Weight loss

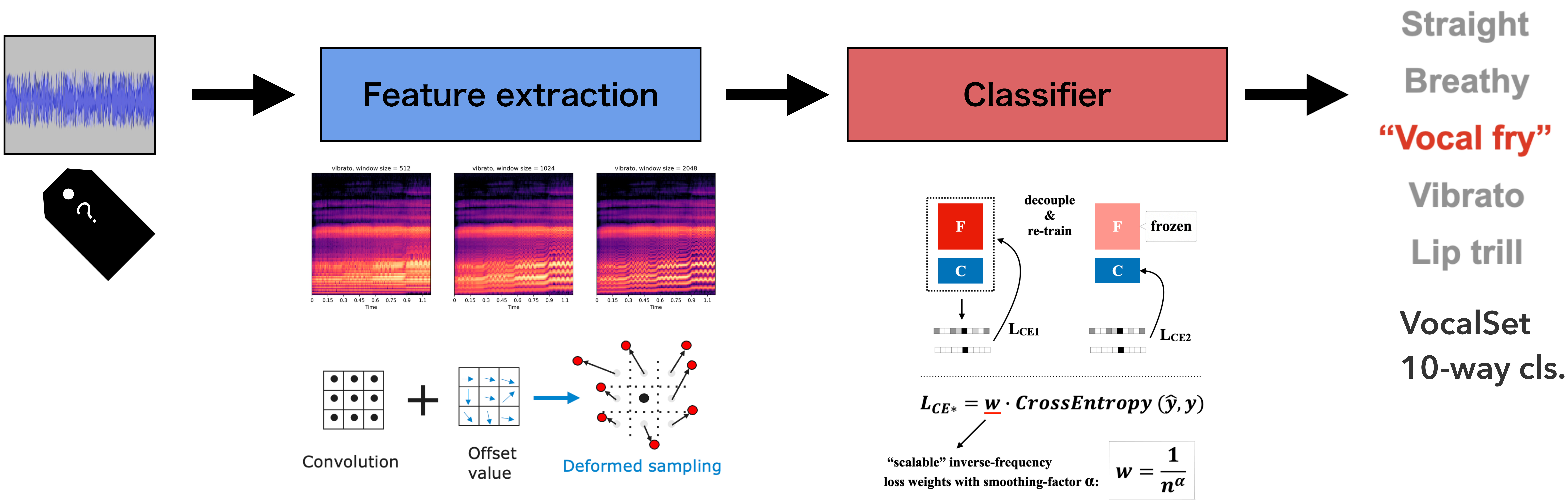
+ Deformable convolution (On 3rd and 4th layer)

+ cRT, only applied weight loss on retraining phase

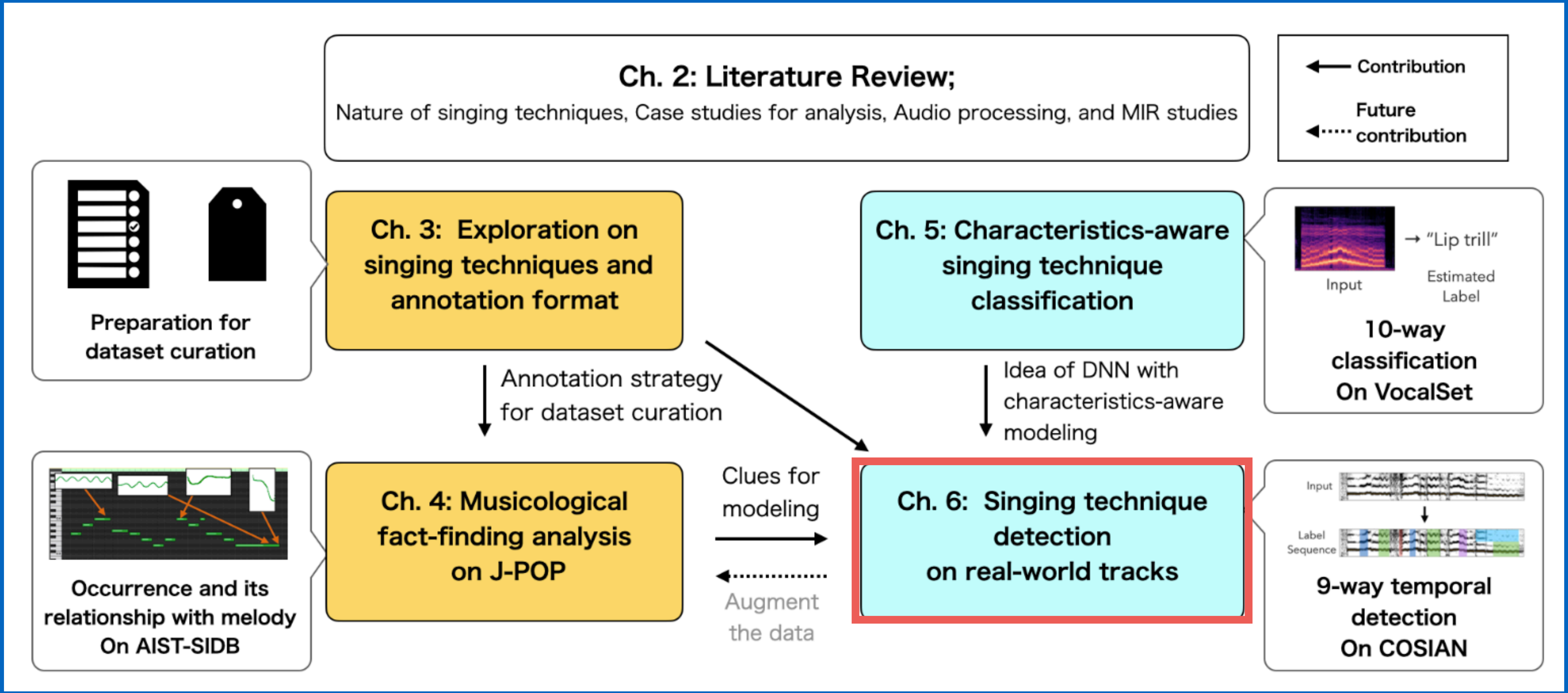
Methods	F1-score	Accuracy	Balanced Accuracy
Multi-resolution Oblong CNN	0.404	0.492	0.472
Multi-resolution Oblong CNN w/ weight loss (α=0.2)	0.513	0.554	0.575
Deformable CNN w/ weight loss (α=0.2)	0.559	0.610	0.635
Deformable CNN w/ weight loss on cRT (α=0.2)	**0.620**	**0.656**	**0.655**

Summary of Chapter 5

Characteristics-aware CNN model for singing technique classification



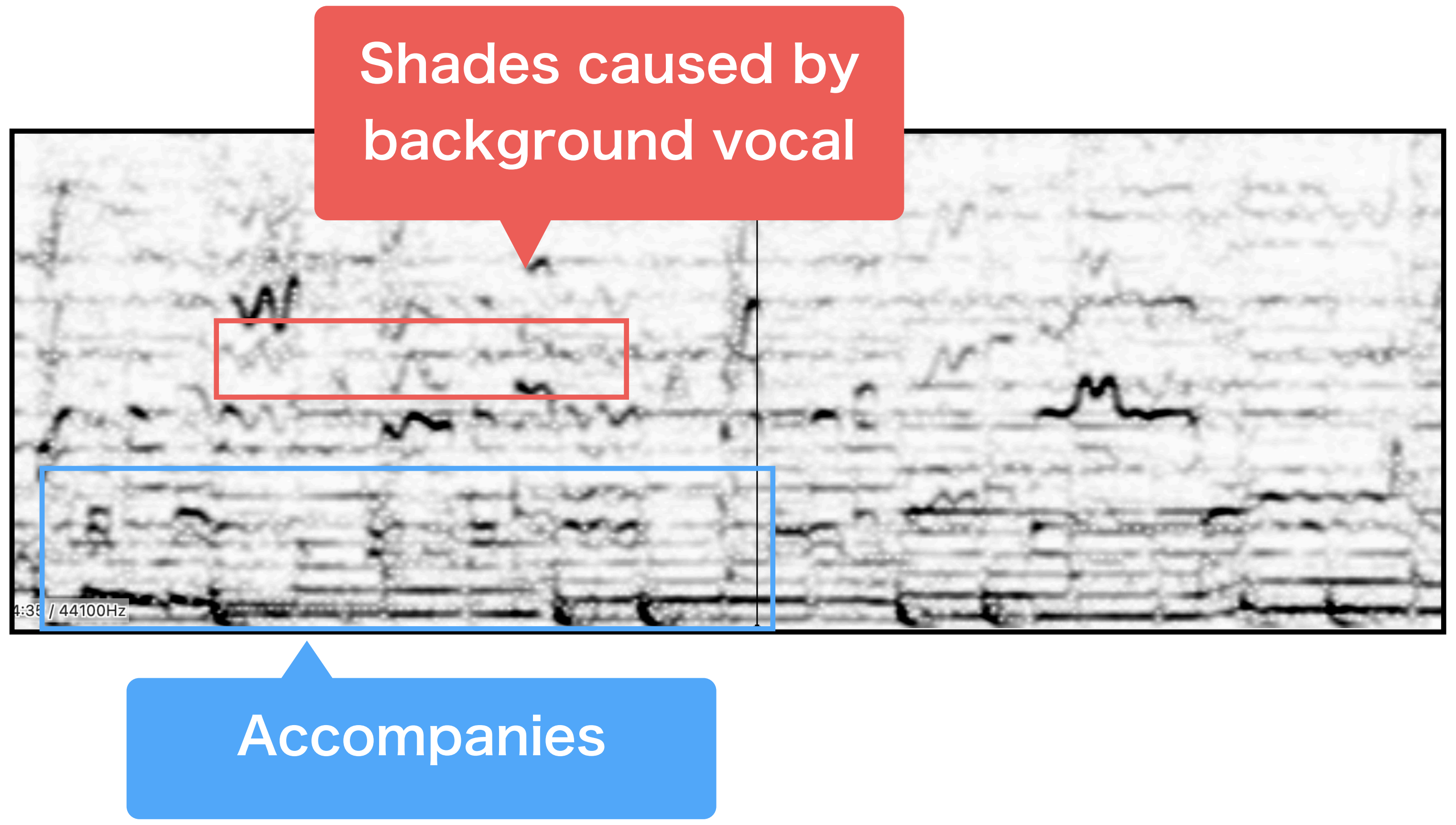
- Contributions:
  - Confirmed the effectiveness of **Multi-resolution** spectrogram and **CNN-kernel modification**
  - Proposed the model based on **Deformable CNN** and **Imbalance-aware learning**



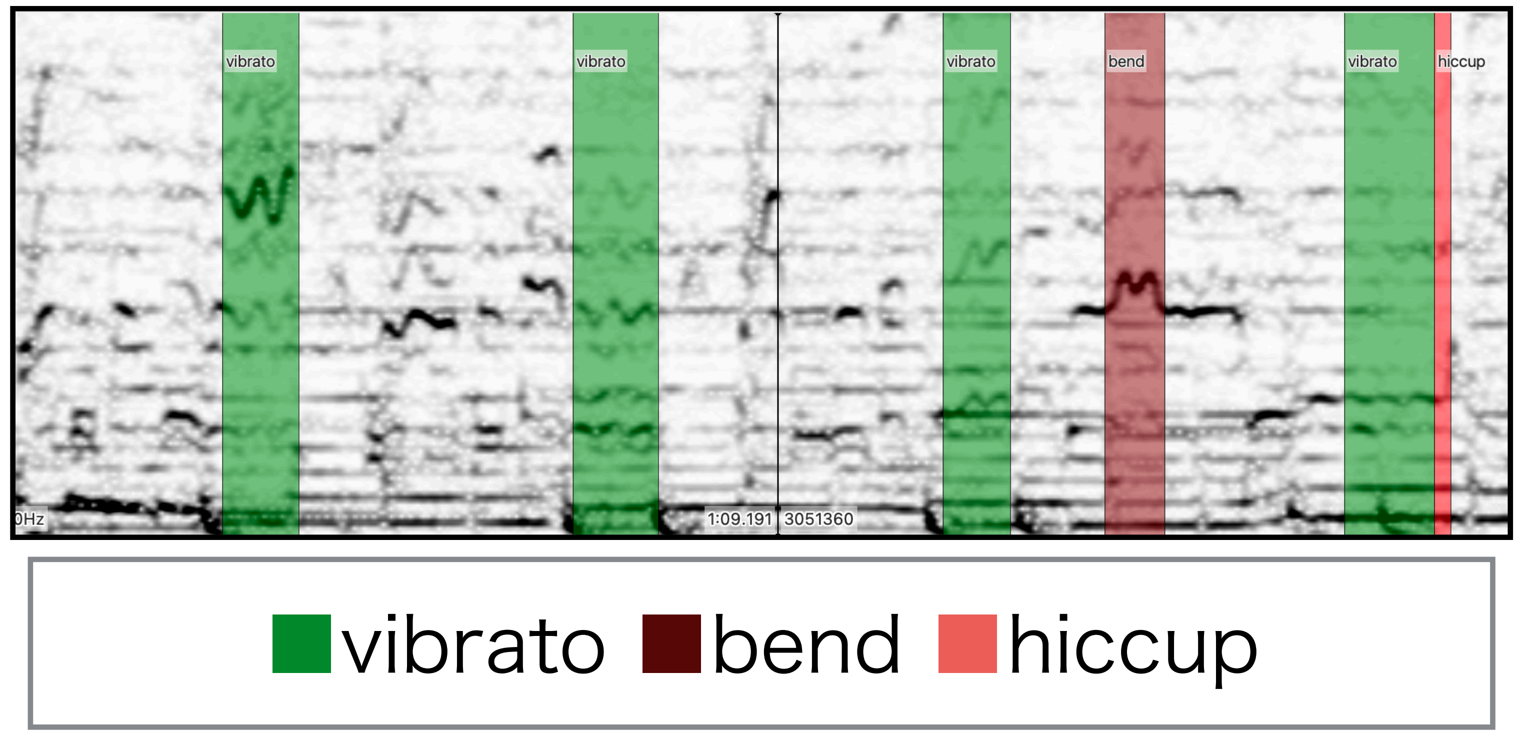
Chapter 6
Singing Technique Detection from Real-world Popular Music

Summary of Chapter 6

Extracting appearance of singing techniques on real-world vocal tracks



Singing voices from commercial musics
(i.e., convolved effects & backgrounds)

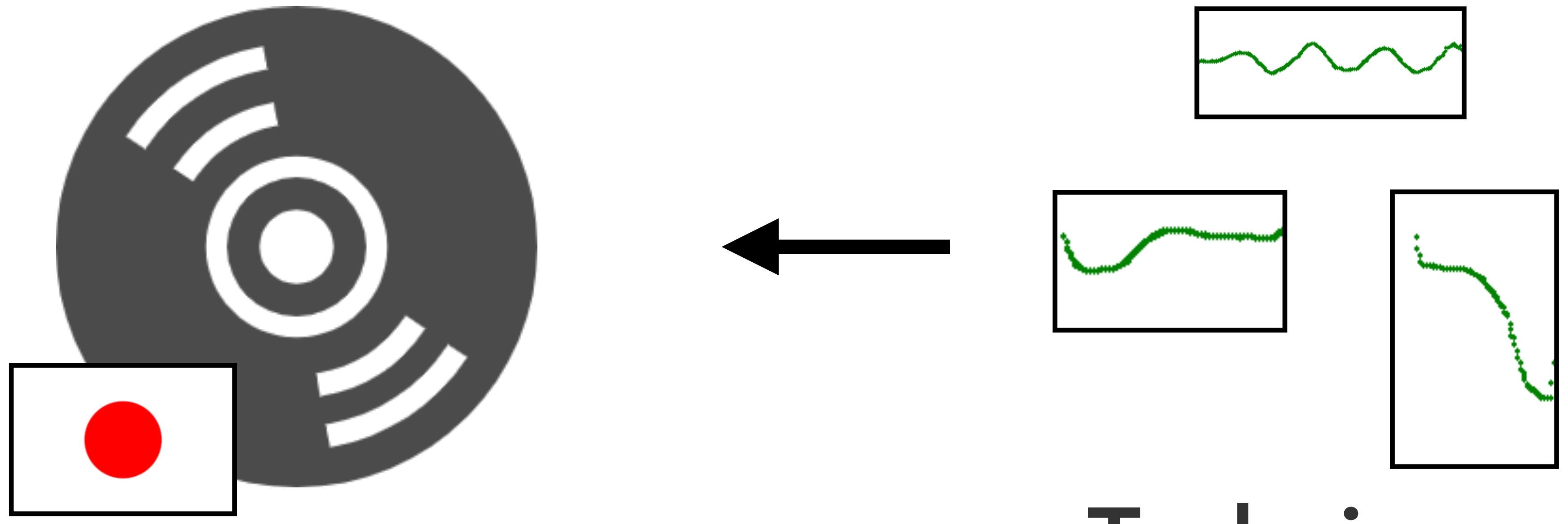


Goal: Automatically detects singing technique appearance

- Challenges
  - Task/Dataset creation -> How to implement a computational task?
  - Detection model -> How to handle temporarily appearing techniques and real-world tracks?

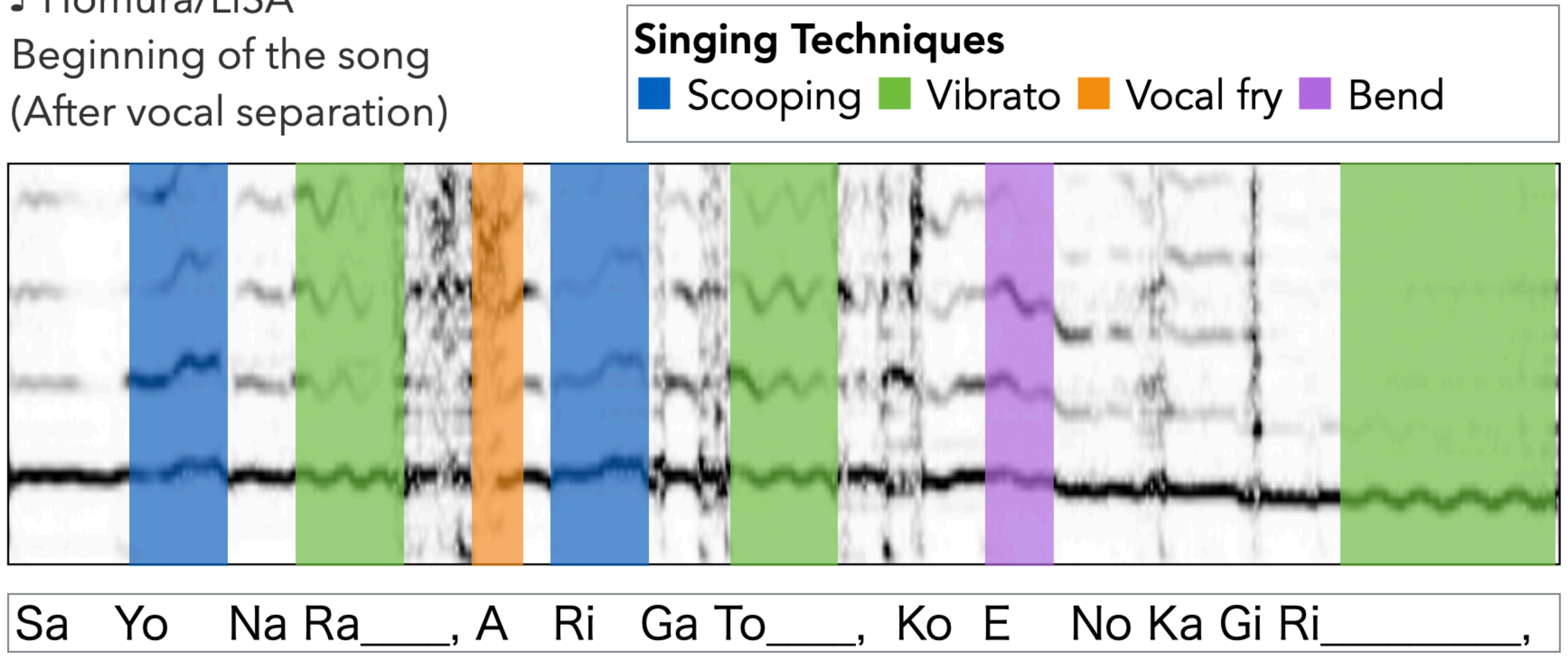
COSIAN: A new dataset

Built a dataset to enable training model and evaluation



Techniques

♪ Homura/LiSA
Beginning of the song
(After vocal separation)

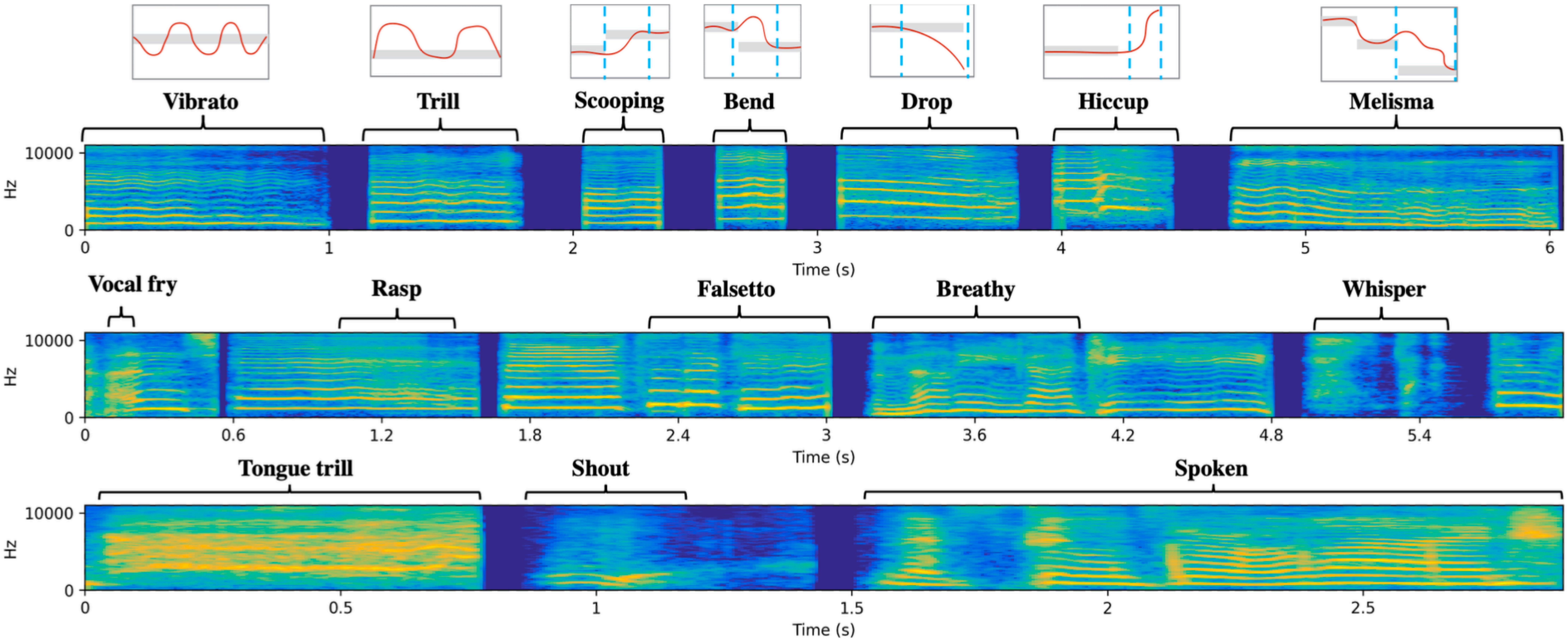


- <u>168</u> **commercial J-POP songs' first section**
  - 21 male, 21 female, various types
  - 4 songs from each vocalist
  - **4h 47m 39s**, in total
  - Voice is separated by Demucs v3 [Defossez 21]

- **Annotation**
  - Region labels of singing techniques
  - Pitch f0 value

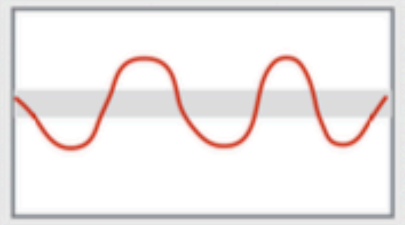
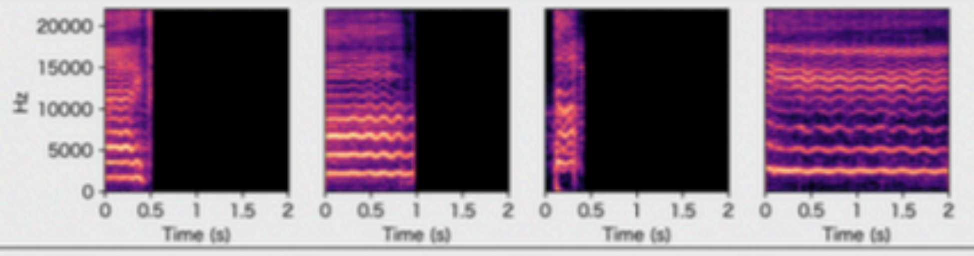
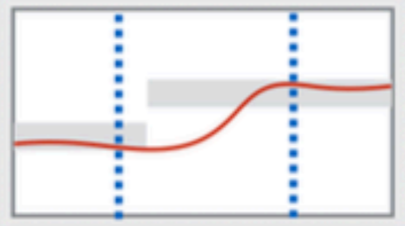
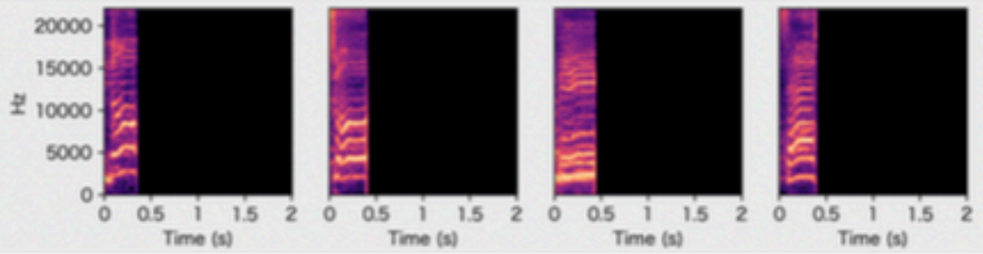
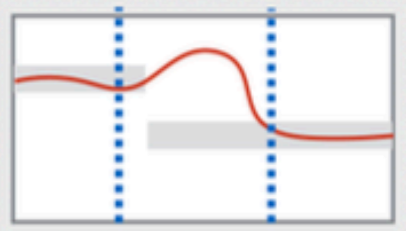
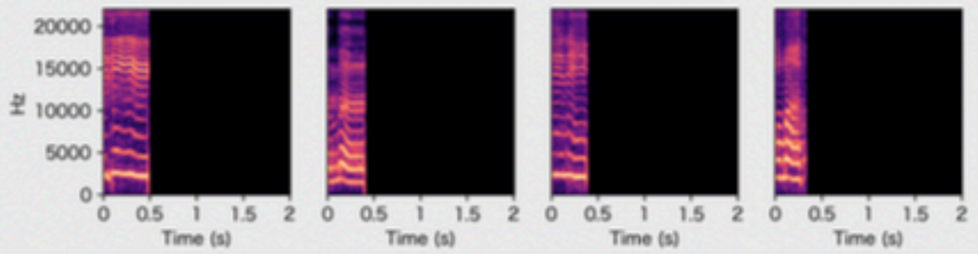
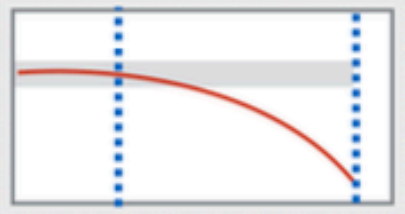
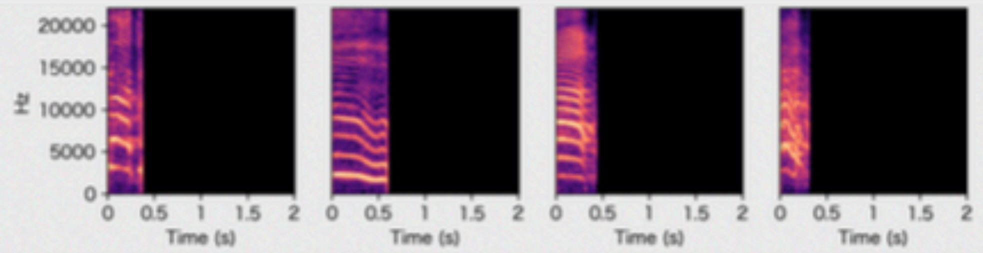
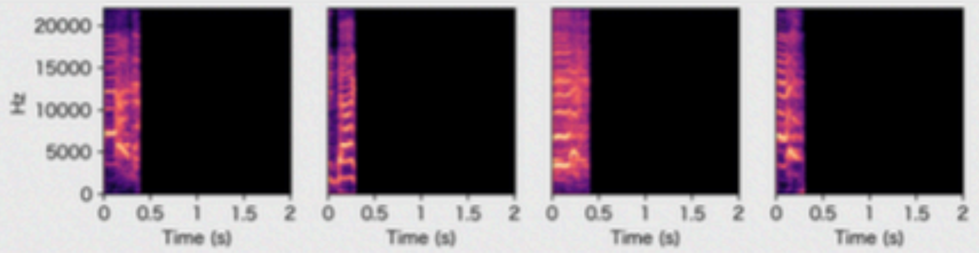
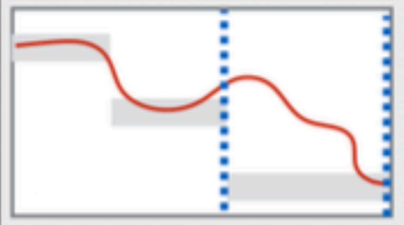
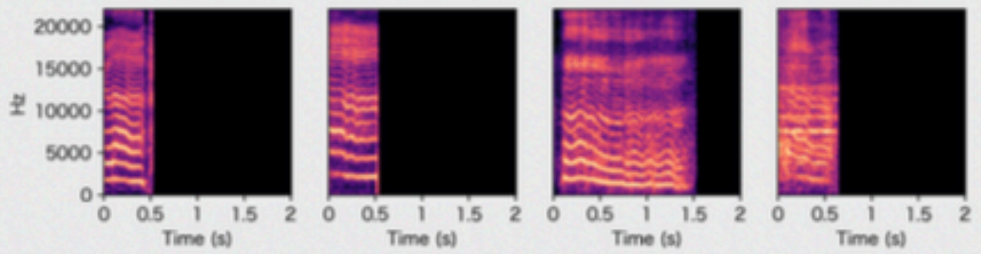
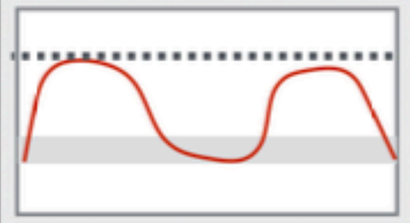
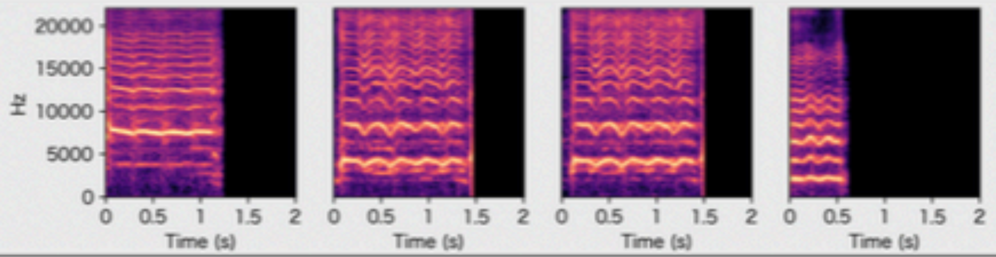
Annotated techniques

Various 15 techniques



Annotated criteria

States the detailed criteria for each techniques (p.69-70)

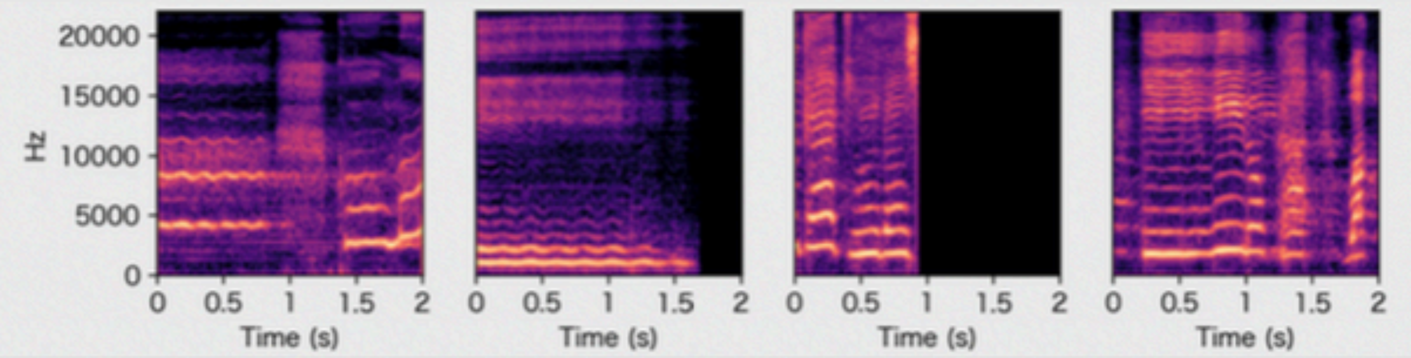
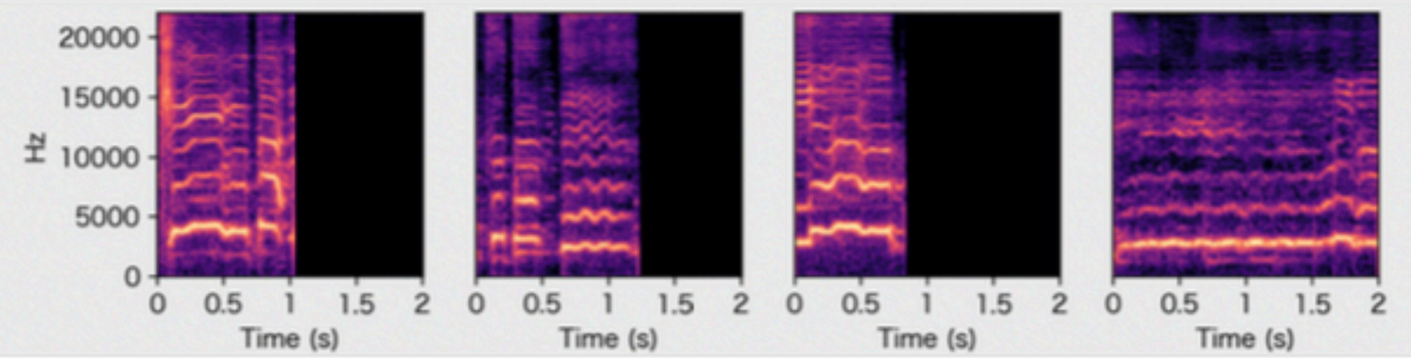
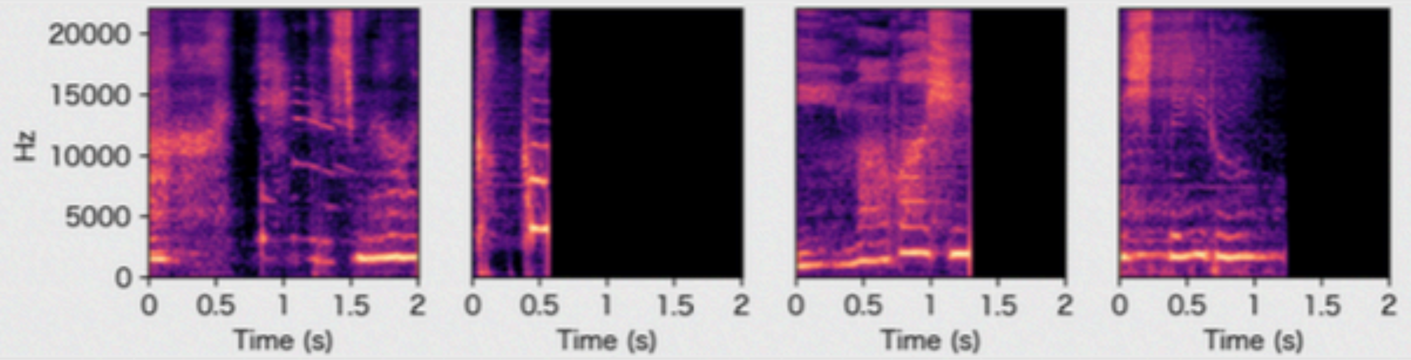
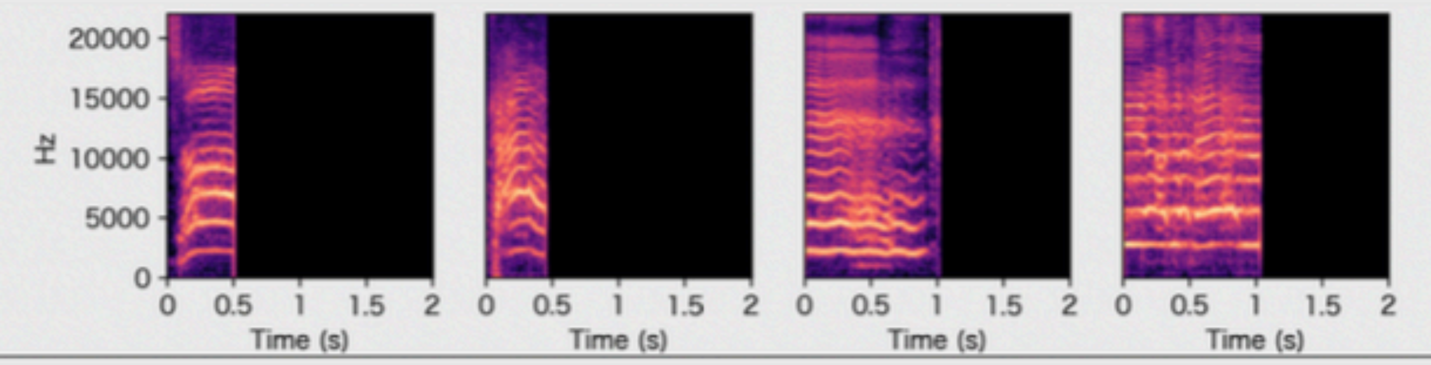
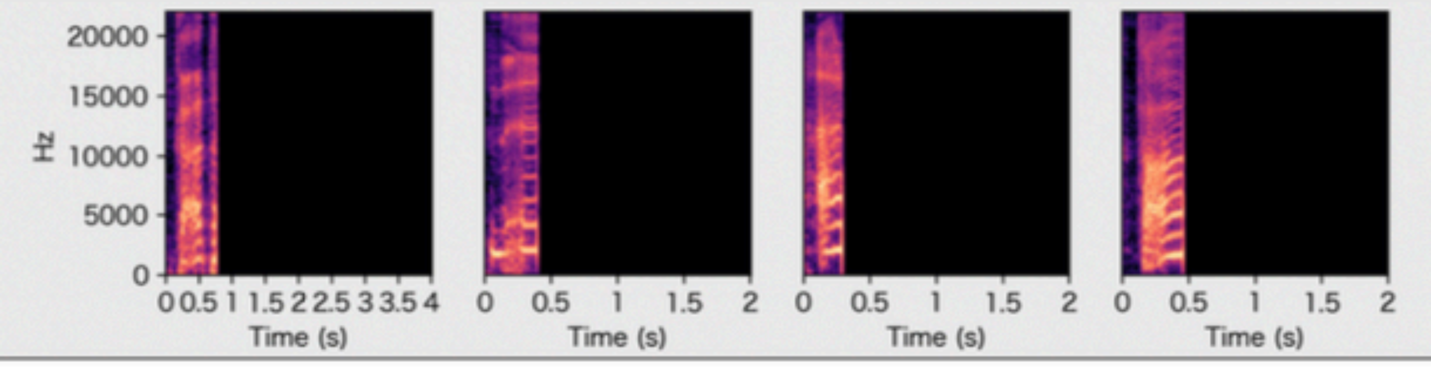
Technique	Sketch	Beginning	End	Difference with...	Samples from audio (mel spectrogram)
Vibrato		Visible beginning of the pitch change	Visible end of the pitch change	**w/ NA**: has visible sinusoid and periodicity **w/ Trill**: does not have the target pitch of the edge of pitch endpoint	
Scooping		Visible beginning of the preparation	Visible end of the overshoot	**w/ NA**: has hearable pitch change **w/ Hiccup**: occurs on the attack and does not have abrupt higher pitch change	
Bend		Visible beginning of the preparation	Visible end of the unstable pitch	**w/ NA**: has hearable pitch change **w/ Vibrato**: <1 roundtrip of the pitch **w/ Hiccup**: not so abrupt pitch change	
Drop		Visible beginning of the pitch dropping	Visible end of the pitch dropping	**w/Bend**: occurs on the release	
Hiccup		Visible beginning of pitch rising	Visible end of pitch region	**w/Bend**: has extreme pitch rising (> 4 semitones) **w/Falsetto**: has instantaneous region of high pitch	
Melisma		Visible beginning of pitch change	Visible beginning of stable pitch region	**w/ NA**: only has one syllable and fast pitch change **w/ Bend**: > 1 stable pitch targets	
Trill		Visible beginning of pitch change	Visible end of pitch change	**w/ NA**: onley has one syllable **w/Vibrato**: has the target pitch of the edge of pitch endpoint	

For pitch: p.69

For timbre: p.70

Annotated criteria

States the detailed criteria for each techniques (p.69-70)

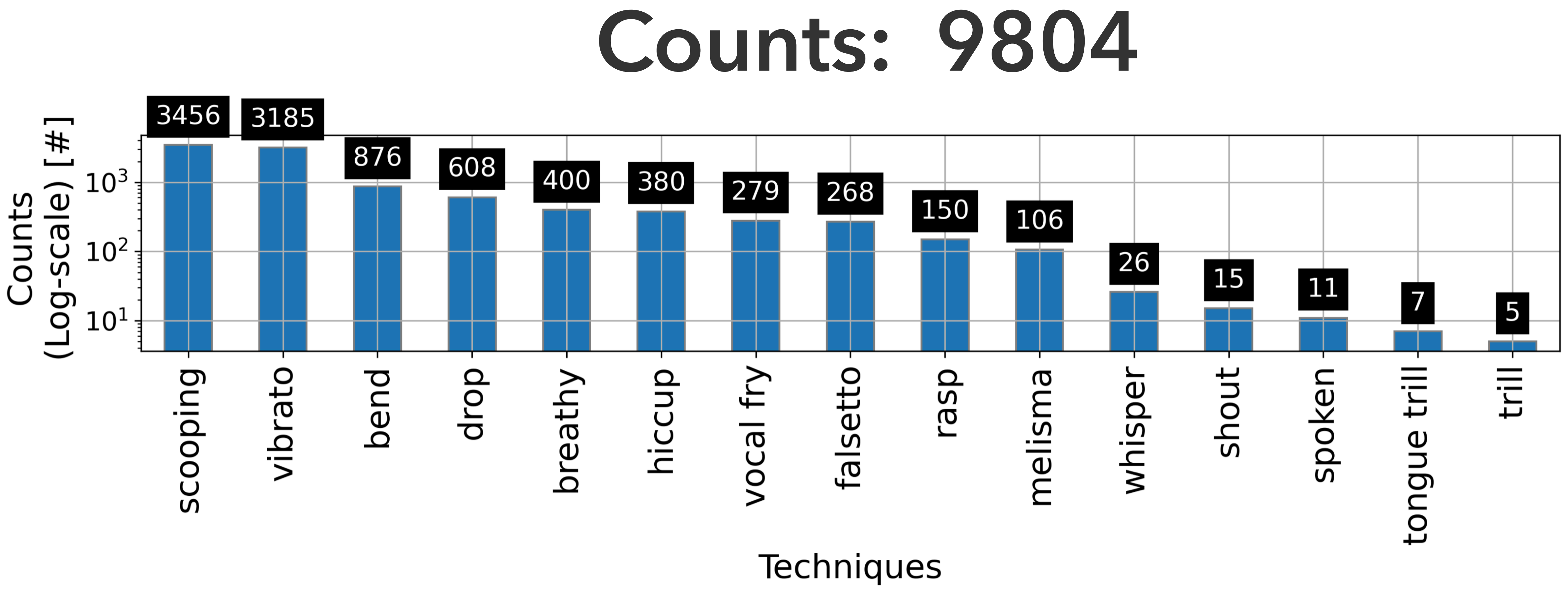
Technique	Difference with…	Samples from audio (mel spectrogram)
Breathy	**w/ NA**: has higher breathiness and more frequential noisy components compared to ordinary voice region **w/ Whisper**: its pitch component is not so missing, relatively **w/ Falsetto**: its vocal register is not falsetto (mixed or modal, etc.)	
Falsetto	**w/ NA**: accompanied by high vocal note and is in different register as ordinary **w/ Breathy**: its vocal register is falsetto **w/ Hiccup**: the region sung by falsetto register is not instantaneous	
Whisper	**w/ Breathy**: its pitch component is relatively missing	
Rasp	**w/ NA**: has distorted timbre, with visible subharmonics on spectrogram **w/ Vocal fry**: has main accompanied pitch	
Vocal fry	**w/ NA**: has creaky sound, with visible pulse pattern on spectrogram **w/ Rasp**: more instantaneous, not accompanied main pitch and tend to be used in attack	

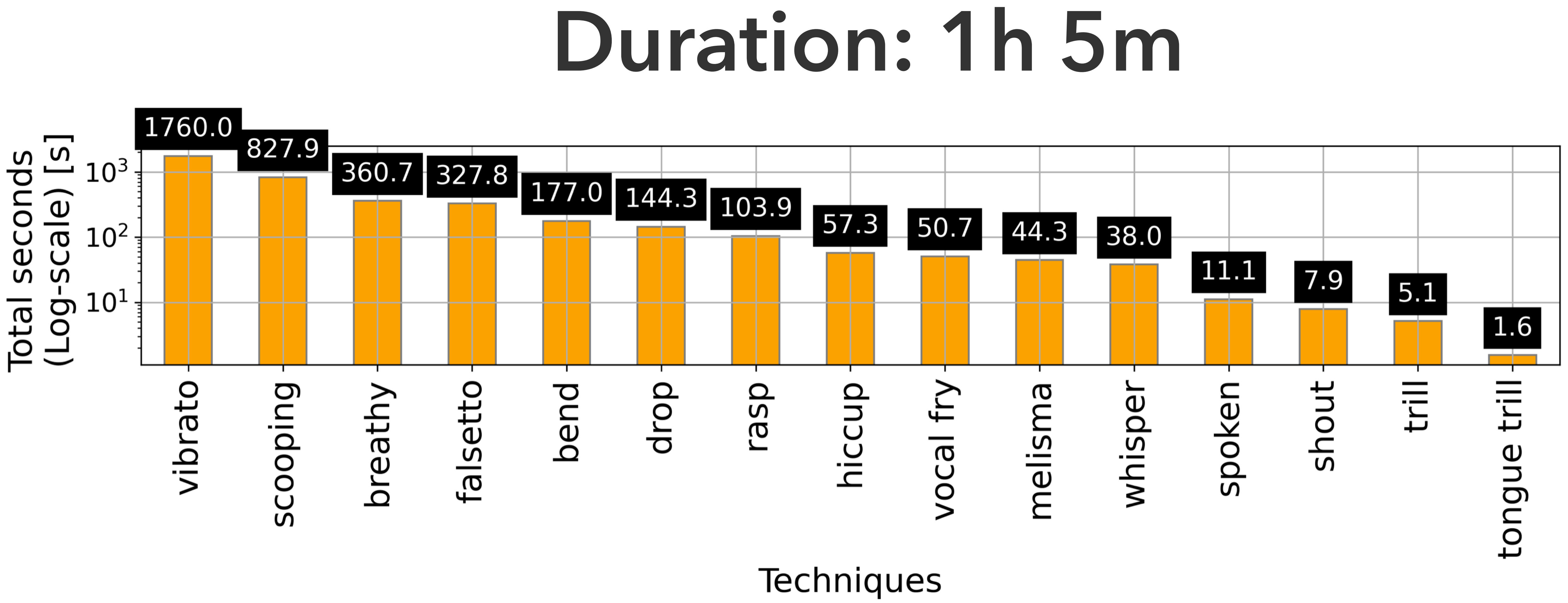
For pitch: p.69

For timbre: p.70

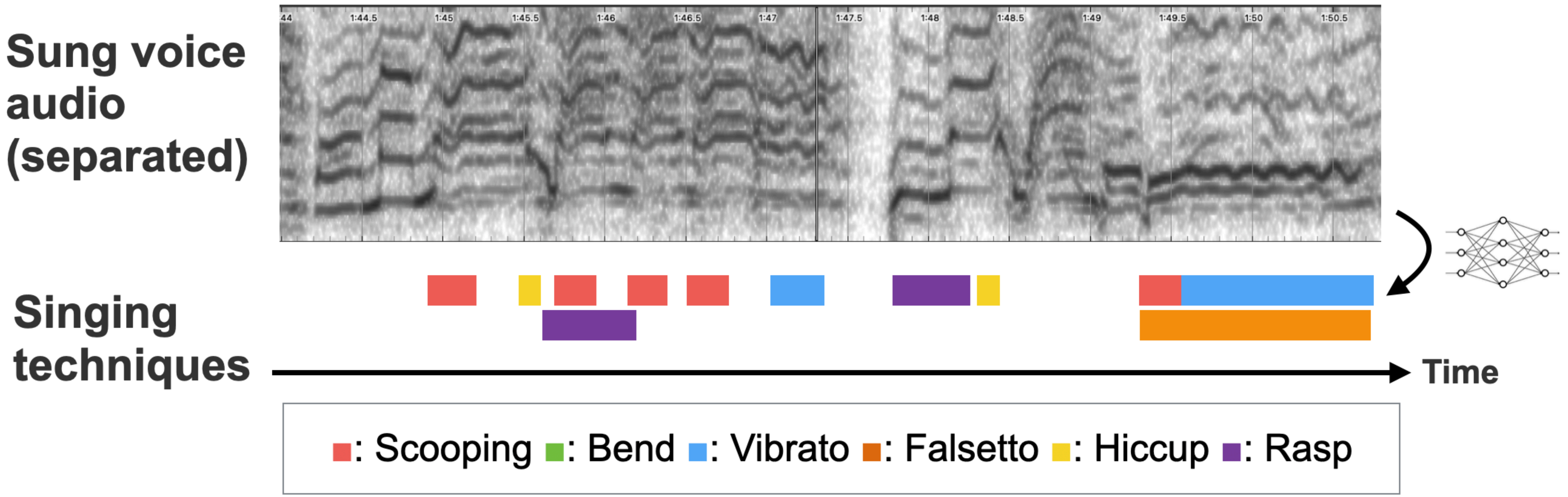
Distribution of labels

- **Technique ratio**
  - vs audio length: **22.8%**
  - vs vocal length: **38.1%**
- **Vibrato and Scooping are most frequent techniques**



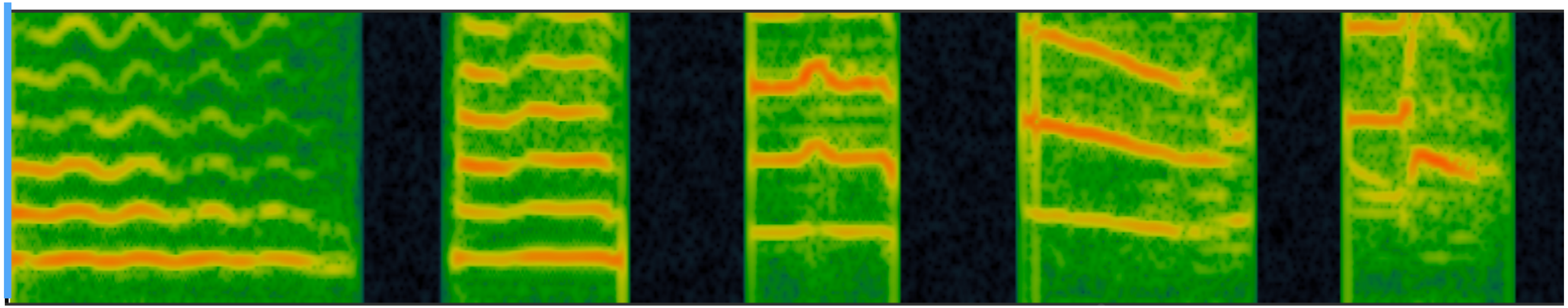


Task: Singing technique detection



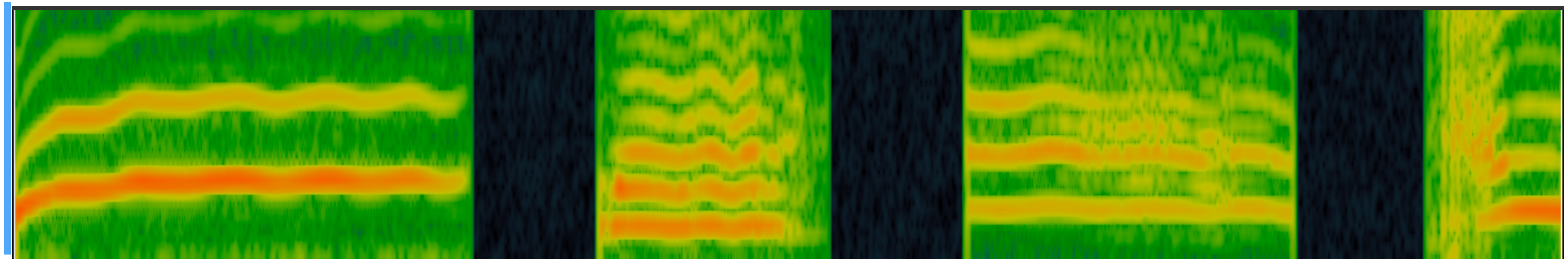
Target: 9 singing techniques

Pitch techniques



Vibrato Scooping Bend Drop Hiccup

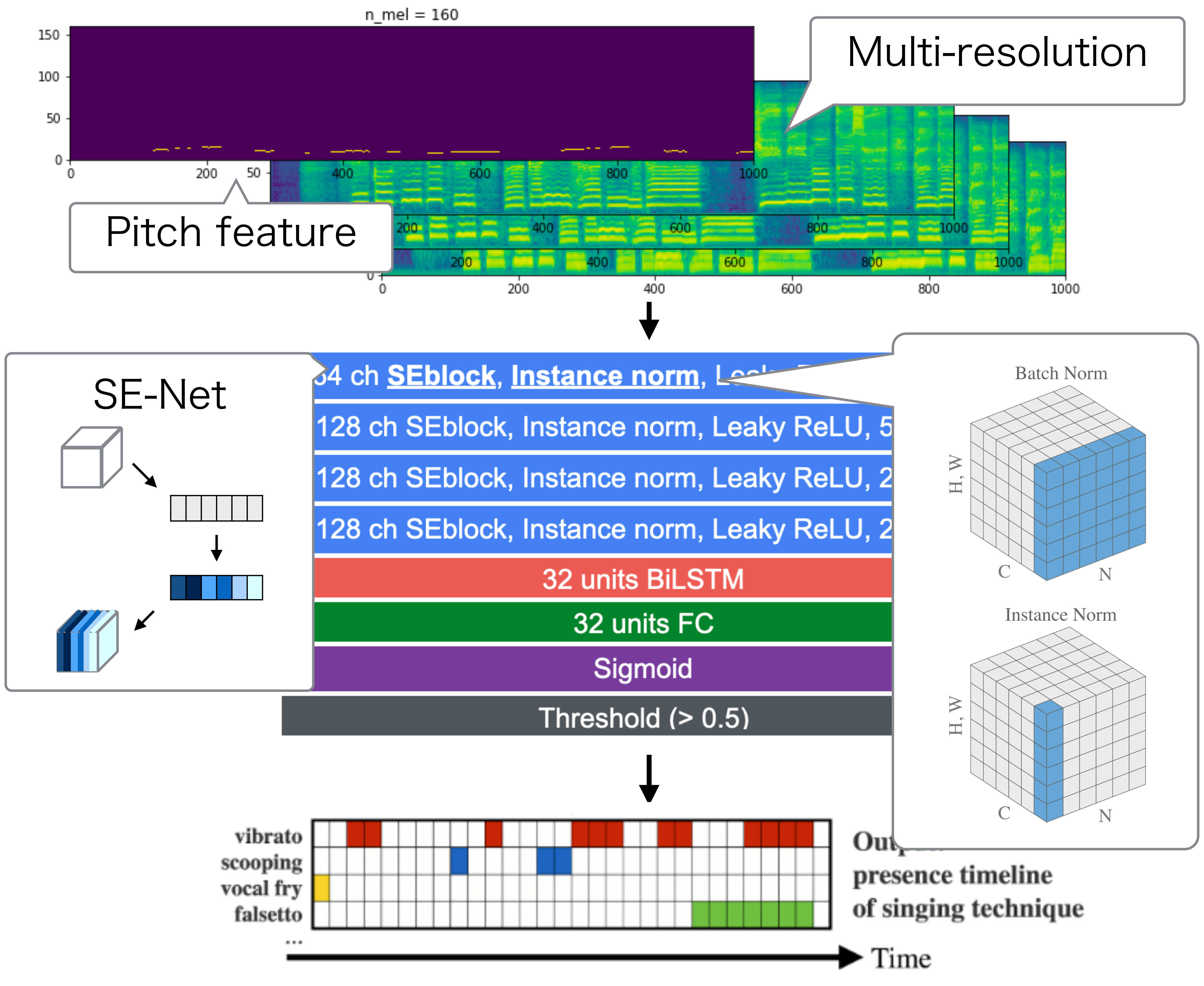
Timbre techniques



Falsetto Breathy Rasp Vocal fry

PrimaDNN': Singing technique detection CRNN

- **Input:**
  - Multi-resolution mel-spectrogram
  - Pitch feature -> estimated from CREPE [Kim 18]
- **Model**
  - CNN based feature extraction
    - Squeeze and Excitation (SE-Net) [Hu 18]
    - Instance normalization
  - RNN (LSTM) based temporal model
- **Loss function**
  - Focal loss [Lin 17] -> empirically set $\alpha=0.13$, $\gamma=1.33$



$$Focal = -\alpha_t(1-p_t)^{\gamma}\log(p_t)$$

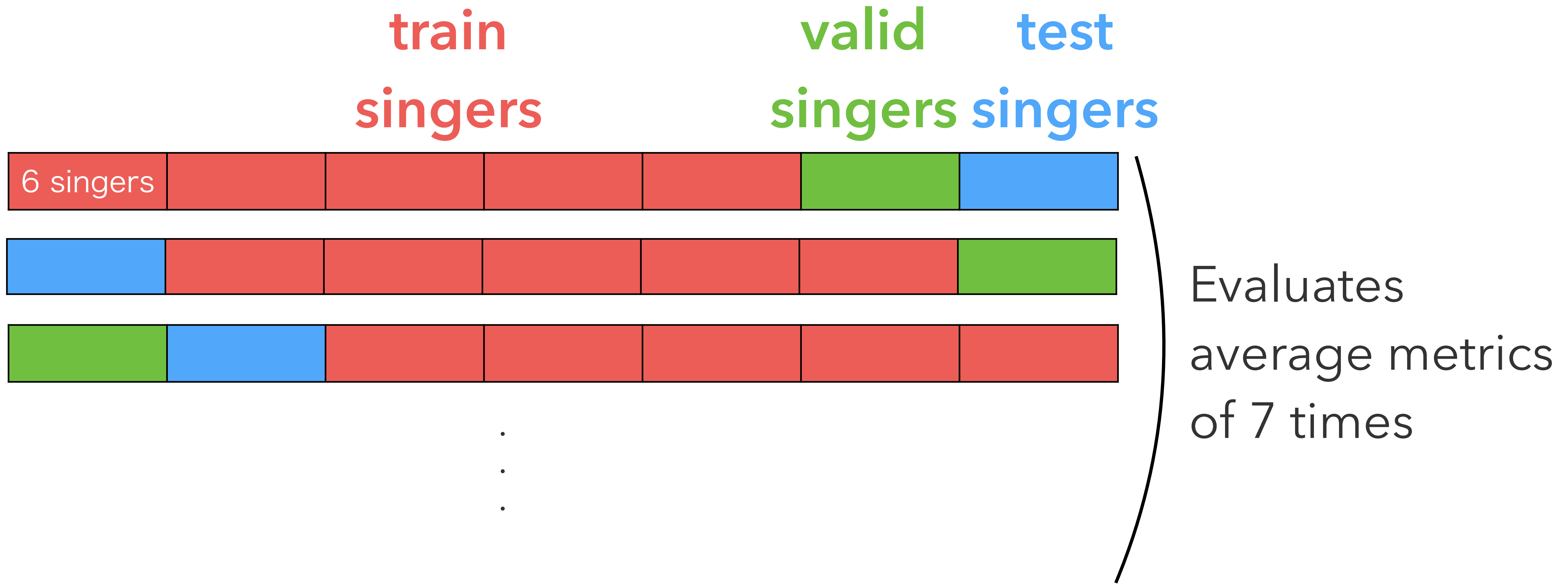
[Kim 18]: J.W.Kim et al. CREPE: A CONVOLUTIONAL REPRESENTATION FOR PITCH ESTIMATION, ICASSP 2018.
[Hu 18] J. Hu et al. Squeeze-and-Excitation Networks. CVPR 2018
[Lin 17] T. Lin et al. Focal loss for Dense Object Detection ICCV 2017

Brief description about each modification of PrimaDNN’

Modifications	Brief explanation	For what?
Pitch feature	Highlights the pitch heights	To take a hint of the melody
Focal loss	Up-weights the hard-samples, Applied for imbalance problems	To enhance the detection of technique segments (sparse)
Multi-resolution	Stacking three different resolutional Mel-spectrograms	To capture a wide type of fluctuation (Similar to Chapter 5)
Squeeze & Excitation (SE)	Weights the importance of feature maps along channel axis	To pick up the more important feature maps -> adapts variation more
Instance Normalization	Calculates normalization at instance level, not batch level	To suppress the effects from Non-targets (singer, effect, etc)

Experimental conditions

- **Singer-wise 7-fold cross validation**
  - In order to evaluate unseen situation
  - Validation set are used for controlling training time (early stopping)

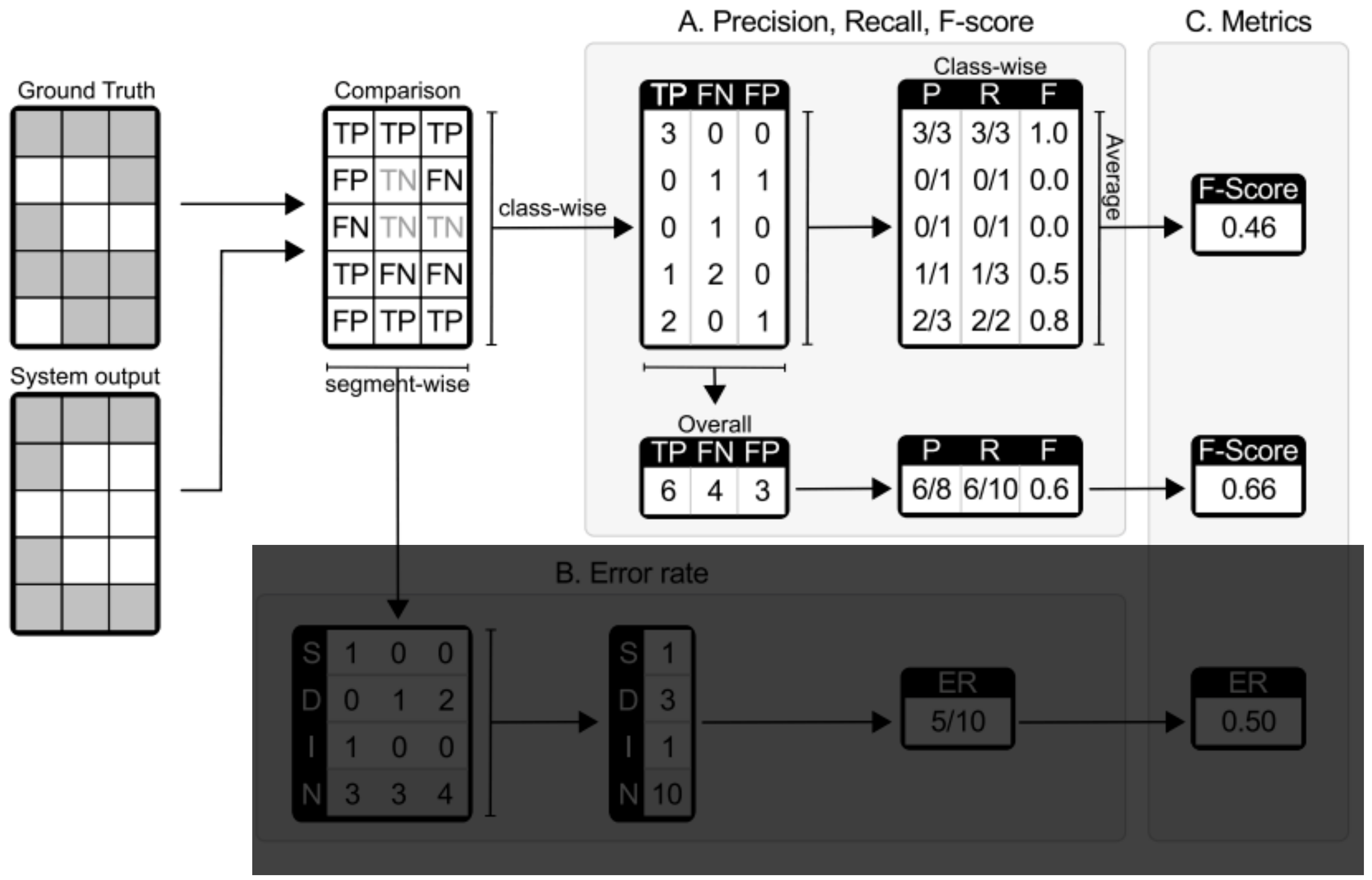


- **Evaluation: Segment-based metrics**
  - Precision (P), Recall (R)
  - **Macro-F**: class-wise average of F-measure
    - Equally consider every classes
  - Micro-F: overall average of F-measure
    - Emphasis more on majority classes
  - Segment length: 50 [ms]

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F = \frac{2 \cdot P \cdot R}{P + R}$$



Metrics for Polyphonic Sound Event Detection,
T. Heittola et al., Applied Science 2016

Comparison

PrimaDNN' achieved the best performance

Hand-crafted feature to DNN

+ Pitch feature & Focal loss

+ Scale-up & Multi-res spec. & SE-Net & Instance norm.

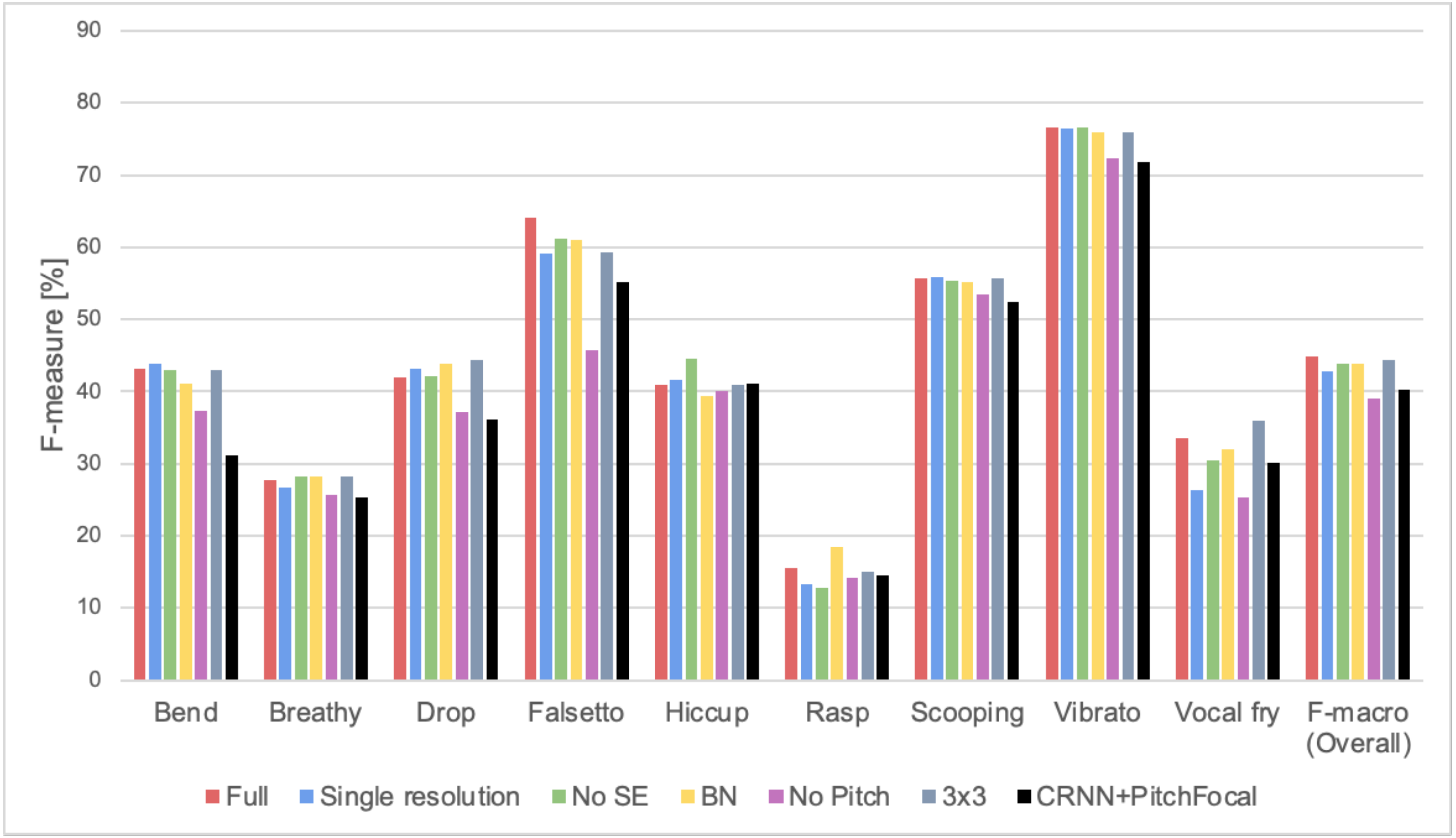
Methods	Macro-F	Micro-F	Precision	Recall
eGeMAPS+LSTM [Eyben 15]	0.092	0.063	0.113	0.016
CRNN (versatile) [Imoto 21]	0.377	0.563	0.422	0.392
CRNN-PitchFocal [ISMIR 22] (ours)	0.402	0.551	0.377	0.480
PrimaDNN' [EUSIPCO 23] (ours)	**0.449**	**0.606**	**0.438**	**0.483**
CNN Self-attention [Imoto 21]	0.420	0.593	0.434	0.477

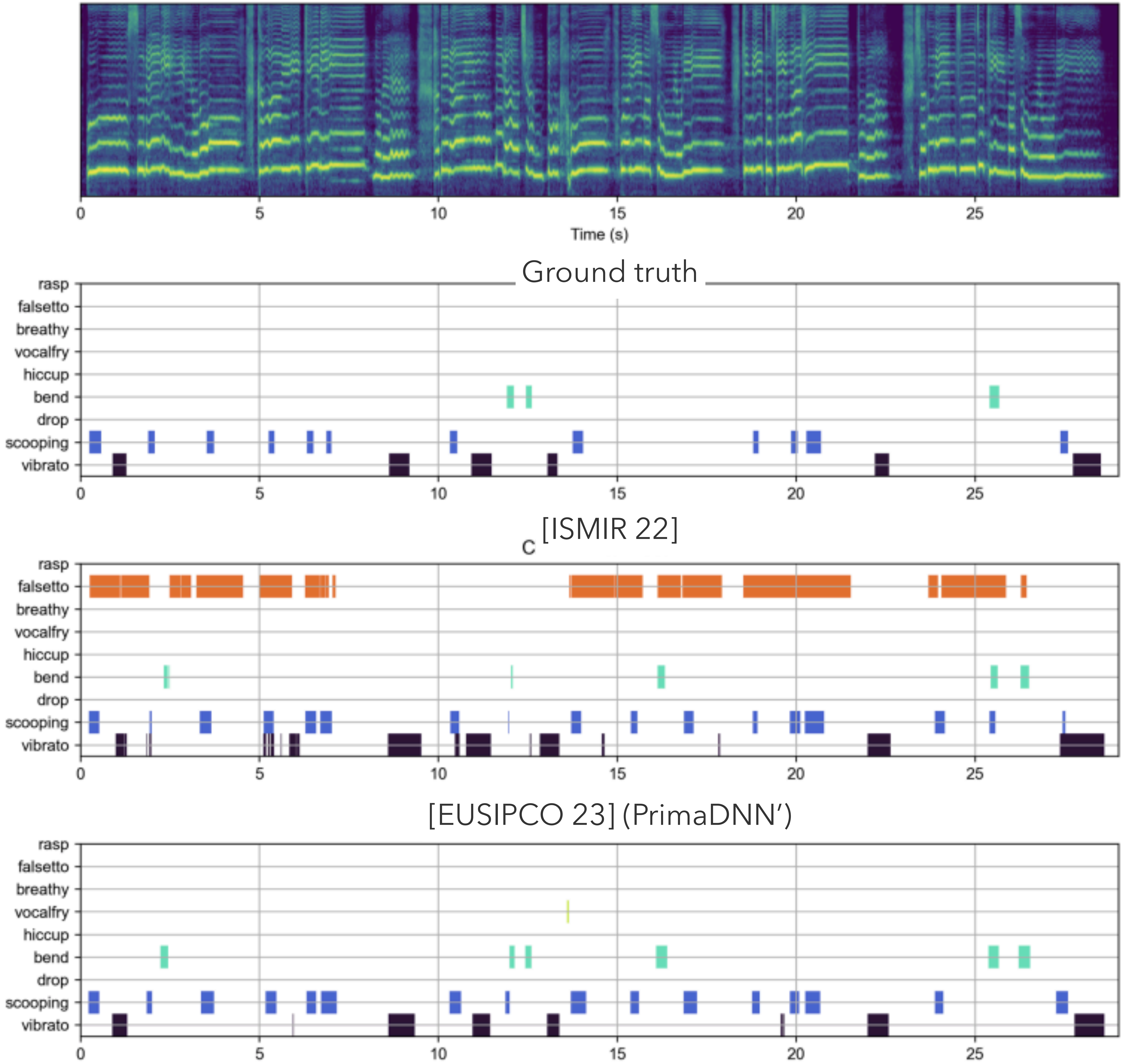
[Eyben 15] F.Eyben et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing, IEEE Trans. affective computing 2015
[Imoto 21] K. Imoto et al. Impact of Sound Duration and Inactive frames on Sound Event Detection, ICASSP 2021
[ISMIR 22] Y. Yamamoto et al. Analysis and Detection of Singing Techniques in Repertoires of J-POP Solo Singers, ISMIR 2022
[EUSIPCO 23] Y.Yamamoto et al. PrimaDNN': A Characteristics-aware DNN Customization for Singing Technique Detection, EUSIPCO 2023

Ablation study

Methods	Macro-F	Micro-F	Precision	Recall
Full model	**0.449**	**0.606**	0.438	0.483
w/o pitch	0.390	0.548	0.366	0.473
w/o multi-resolution (Only use single resolution)	0.429	0.602	0.441	0.466
W/o Squeeze & Excitation	0.438	0.603	0.430	0.481
W/o Instance norm. (Use batch norm)	0.439	0.596	**0.446**	0.481
W/o wide kernel (3x3 kernel)	0.443	0.600	0.432	**0.488**

Ablation study





♫ Yo Hitoto, Hanamizuki

Notable improvements

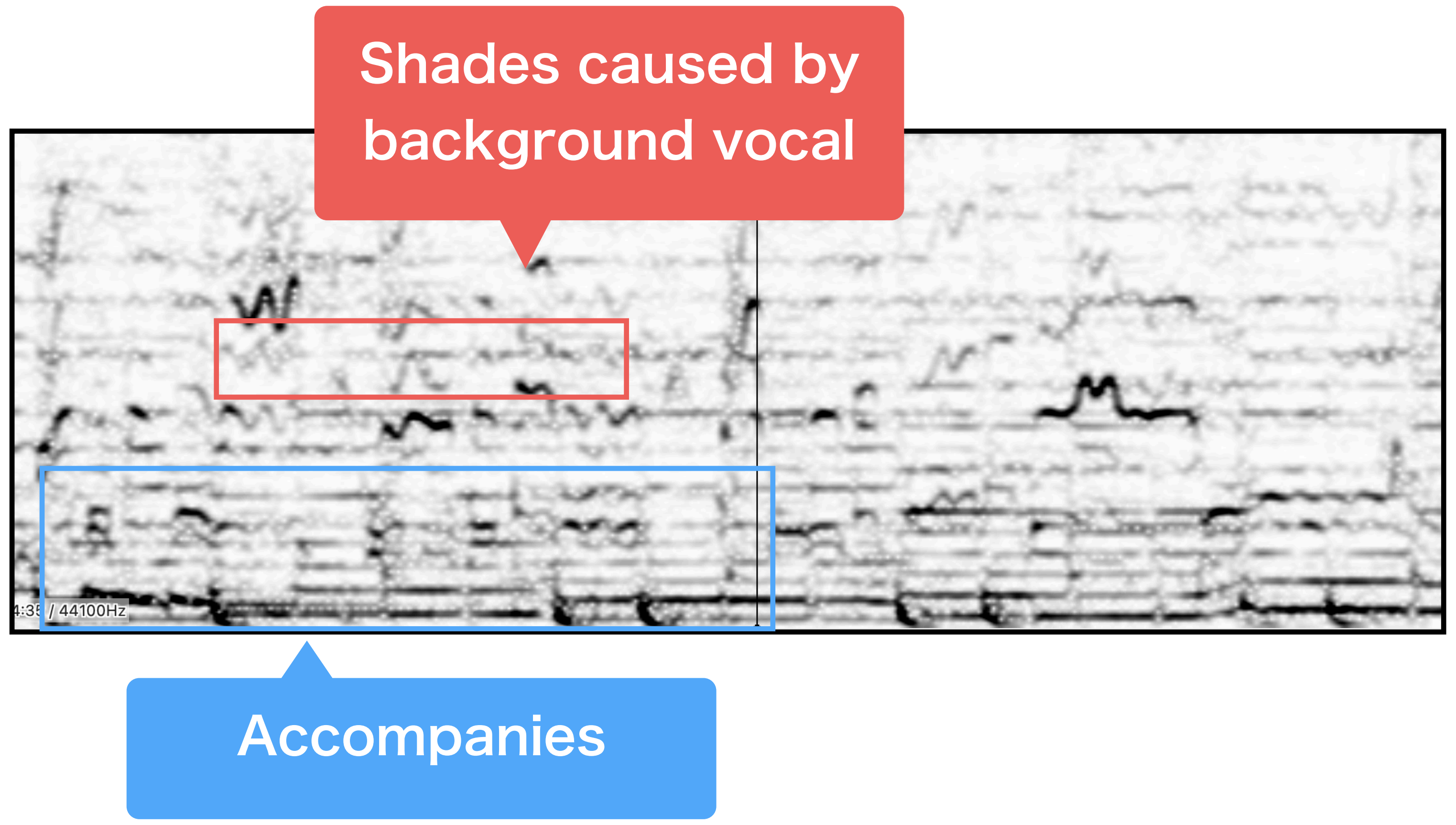
1: CRNN+PitchFocal misidentified falsetto, which is not appeared in the track
while PrimaDNN′ accurately identifies.

2: PrimaDNN′ is more robust on the localization of pitch techniques

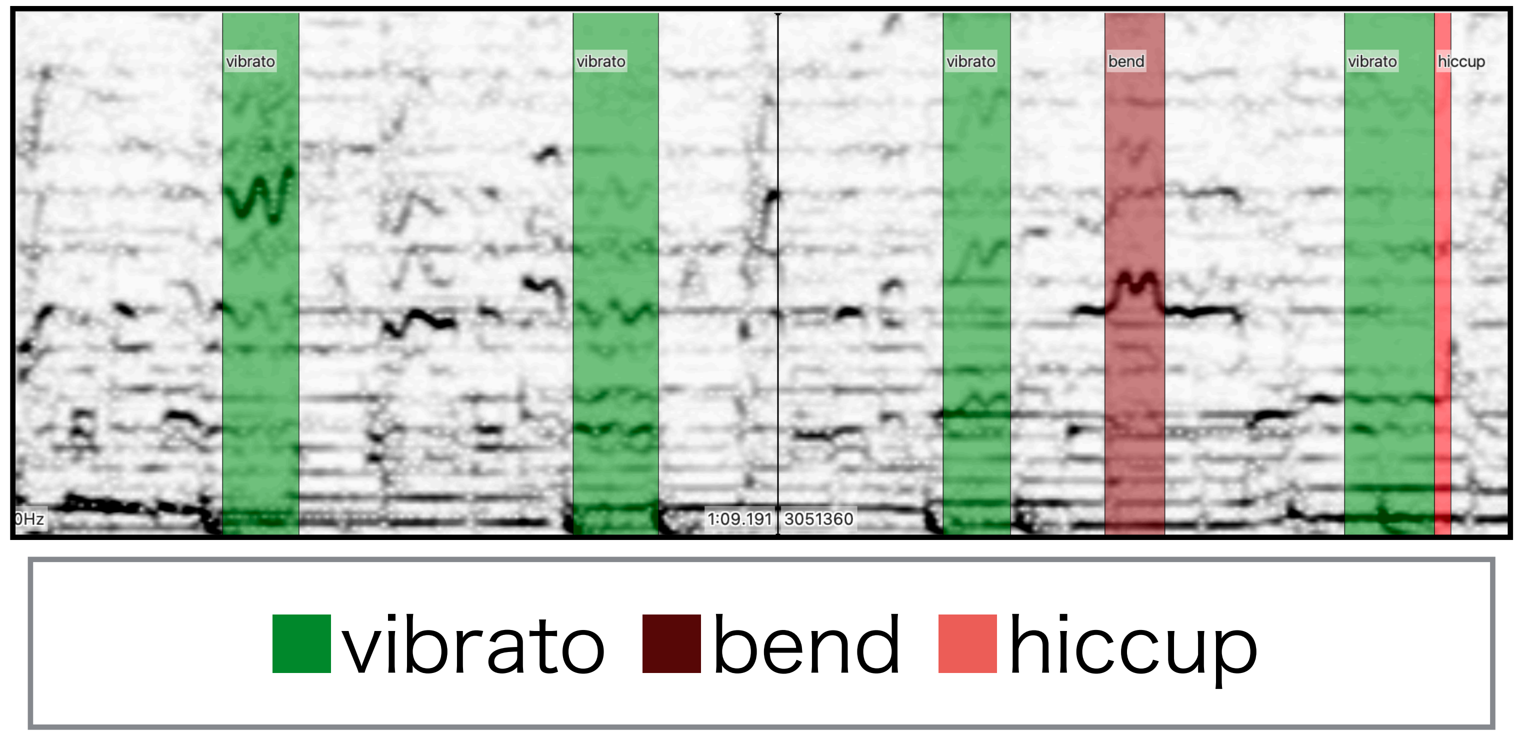
←More example

Summary of Chapter 6

Extracting appearance of singing techniques on real-world vocal tracks



Singing voices from commercial musics
(i.e., convolved effects & backgrounds)



Goal: Automatically detects singing technique appearance

- **Contributions**
  - **Build a new MIR task with the first temporarily annotated datasets on J-POP tracks**
  - **PrimaDNN', A SoTA model for 9-way singing technique detection considering characteristics of data (i.e., Input modification, CNN modification, and Focal loss)**

Conclusion

- **1. Absence of data and its characteristics**
  - What singing techniques should be annotated?
  - Need datasets, but how to annotate?
  - What is their specific characteristics?

Summary of Techniques (Ch.3)

Adopt region labels (Ch.3)
→Demonstrated analysis (Ch.4)

Find some discoveries (Ch.4)

- **2. Less established identification methods**
  - How to design/evaluate the automatic model?
  - Can we detect techniques from real-world singing tracks?

Modified DNN with spectrogram input seems better (Ch.5/6)
Data imbalance-aware training is effective (Ch.5/6)

Built a dataset to evaluate it (Ch.6)
Nine-common techniques can be detected by customized model at 44.9% on Macro-F(Ch.6)

Limitation and future works ①: About data

- **Dataset expansion**
  - Data amount
    - More data and annotation is needed
  - Data quality
    - Associating with other data format (parameter, note-wise annotation, text, etc.)
    - Annotating other information such as vocal notes, lyrics, etc.

Limitation and future works ②: About computation

- **Further improvement on detection**
  - Transfer learning (e.g., self-supervised speech models)
  - Auxiliary task (e.g., onset detection, activity detection, etc.)

- **The range of detection/scope of techniques**
  - There are still undetectable techniques -> Few-shot learning, Anomaly detection, etc.
  - Wider genres -> Cross genre study

- **Association with other application**
  - Combine with singing transcription or other tasks
  - Leverage the detection results for singing voice analysis, synthesis, etc.

Conclusion remarks

Establishment of computational foundation for singing technique analysis

- **Chapter3:** Summarized various singing techniques and set the annotation strategy (i.e., annotate observable techniques by region label) to create datasets
- **Chapter 4:** Conducted singing technique analysis using annotation to investigate the relationship with song and singer.
- **Chapter 5:** Explored singing technique classification models and Proposed DNN models based on characteristics-aware feature extraction and imbalance-aware learning
- **Chapter 6:** Established nine-way singing technique detection on real-world pieces with a new dataset and DNN models with characteristics-aware customization

Achievements and Acknowledgements

Publications related to the thesis

- **Core papers**
  - [IPSJ 23] Yuya Yamamoto, Tomoyasu Nakano, Masataka Goto, Hiroko Terasawa. **Singing technique analysis with correspondence to musical score on imitative singing of popular music.** IPSJ Journal Vol. 64, No.10, 2023 (in Japanese), (Referred, Prized IPSJ Journal Specially Selected Paper (Equal to top-10%)) **Core paper 1**
  - [ISMIR 22] Yuya Yamamoto, Juhan Nam, Hiroko Terasawa. **Analysis and Detection of Singing Techniques in Repertoires of J-POP Solo Singers.** In Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR), 2022 (Referred, acceptance rate: 43%, oral and poster) **Core paper 2**

- **Others**
  - [APSIPA 21] Yuya Yamamoto, Juhan Nam, Hiroko Terasawa, Yuzuru Hiraga. **Investigating Time- Frequency Representations for Audio Feature Extraction in Singing Technique Classification**, In Proceedings of the 2021 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2021 (Referred, poster)
  - [INTERSPEECH 22] Yuya Yamamoto, Juhan Nam, Hiroko Terasawa. **Deformable CNN and Imbalance-aware Feature Learning for Singing Technique Classification**. In Proceedings of the 23rd Annual Conference of the International Speech Communication Association (INTERSPEECH), 2022 (Referred, acceptance rate: 51%, oral)
  - [EUSIPCO 23] Yuya Yamamoto, Juhan Nam, Hiroko Terasawa. **PrimaDNN': A Characteristics-aware DNN Customization for Singing Technique Detection.** Proceedings of the 31st European Signal Processing Conference (EUSIPCO), 2023 (Referred, poster)
  - [SLIS 21] Yuya Yamamoto, **Establishing foundations for automatic singing technique detection**. (in Japanese), Master dissertation, University of Tsukuba, 2021

Publications (Referred)

7 referred papers, including top-conference (ISMIR, INTERSPEECH) and awarded journal paper

- [APSIPA 23] Yuya Yamamoto, **Toward Leveraging Pre-Trained Self-Supervised Frontends for Automatic Singing Voice Understanding Tasks: Three Case Studies**, In Proceedings of the 2021 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2023 (Referred, oral)
- [EUSIPCO 23] Yuya Yamamoto, Juhan Nam, Hiroko Terasawa. **PrimaDNN': A Characteristics-aware DNN Customization for Singing Technique Detection.** Proceedings of the 31st European Signal Processing Conference (EUSIPCO), 2023 (Referred, poster)
- [IPSJ 23] Yuya Yamamoto, Tomoyasu Nakano, Masataka Goto, Hiroko Terasawa. **Singing technique analysis with correspondence to musical score on imitative singing of popular music.** IPSJ Journal Vol. 64, No.10, 2023 (in Japanese), (Referred, Prized IPSJ Journal Specially Selected Paper (Equal to top-10%))
- [ISMIR 22] Yuya Yamamoto, Juhan Nam, Hiroko Terasawa. **Analysis and Detection of Singing Techniques in Repertoires of J-POP Solo Singers.** In Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR), 2022 (Referred, acceptance rate: 43%, oral and poster)
- [APSIPA 22] Yoshiteru Matsumoto, Hiroyoshi Ito, Hiroko Terasawa, Yuya Yamamoto, Yuzuru Hiraga, Masaki Matsubara. **Human-In-The-Loop Chord Progression Generator With Generative Adversarial Network,** In Proceedings of the 2021 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2022 (Referred, oral)
- [INTERSPEECH 22] Yuya Yamamoto, Juhan Nam, Hiroko Terasawa. **Deformable CNN and Imbalance-aware Feature Learning for Singing Technique Classification**. In Proceedings of the 23rd Annual Conference of the International Speech Communication Association (INTERSPEECH), 2022 (Referred, acceptance rate: 51%, oral)
- [APSIPA 21] Yuya Yamamoto, Juhan Nam, Hiroko Terasawa, Yuzuru Hiraga. **Investigating Time- Frequency Representations for Audio Feature Extraction in Singing Technique Classification**, In Proceedings of the 2021 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2021 (Referred, poster)

Social activities

Reviewing on top journal, many guest talks, educational activities, etc.

- **Reviewer**: IEEE/ACM Transaction on Audio, Speech, and Language Processing (IF: 5.4 at 2023): 2023, 2024
- **Guest talk**:
  - Lightning talk on Music Analysis Meetup (MUANA) 2021, 2022, 2023
  - Komagata junior high school (Career development), 2021
  - Guest talk in SIGMUS 136 (report on ISMIR 2022), 2023
  - Tsukuba University of Technology, 2024
- **Educational activity**:
  - Teaching fellow on Music and Acoustic Information Processing, College of Media Arts, Science and Technology, University of Tsukuba, 2021
  - Organizing paper reading meetup of ISMIR 2022, 2023
  - Organizing weekly lecture on Music and audio X Deep learning, 2022 (https://github.com/yamathcy/music-deeplearning-japanese)
  - Curated list of music and audio processing, 2021 (https://github.com/yamathcy/awesome-music-informatics, 150+ stars on 2024 Jan.)

Awards and grants

Won 6 awards and Got 3 grants, during the graduate school

- Awards
  - **IPSJ Journal Specially Selected Paper (equal to top-10%)**, from IPSJ, 2023
  - **Sound Symposium Student Excellence Presentation Award**, from IPSJ SIGMUS and SIGSLP, 2023, as a co-author (First author: Tsugumasa Yutani)
  - **IPSJ Yamashita SIG Research Award (equal to the annual best paper)**, from IPSJ, 2023 paper title: Analysis of frequency, acoustic characteristics, and occurrence location of singing techniques using imitated j-pop singing voice (at SIGMUS 132, 2021.)
  - **Best Presentation Award (Best research),** from IPSJ SIGMUS, 2021
  - **Dean's Award** of University of Tsukuba, 2021
  - **Student Award**, from IPSJ SIGMUS, 2019
- Grants
  - **JST SPRING, tier1 (top-25%)**
  - **Travel Grant of The Telecommunications Advancement Foundation (JPY 190,000)**, 2022
  - **ISMIR student author grant 100 % wavier,** 2022

Acknowledgements

- **Dr. Hiroko Terasawa, Dr. Nobutaka Suzuki, Dr. Hiroyoshi Ito**: Supervisors
- **Dr. Atsushi Toshimori, Dr. Shuichi Moritsugu**: The committee members
- **Dr. Juhan Nam**: Co-advisor and The guest committee
- **Dr. Masataka Goto, Dr. Tomoyasu Nakano**: Collaboration and Mentor for research works
- **Dr. Yuzuru Hiraga**: Supervisor (B.S. and M.S.)
- **The members of LSPC**
- **The members of MACLab@KAIST**
- **Researchers/Students whom I met in the conferences**
- **JST SPRING grant program**: Financial support for research and living expenses
- **Vocalists in the world**
- **My family**
- **Yuya Yamamoto**: Myself

Thank you!!

A Computational Approach to Analysis and Detection of Singing Techniques

January 29th, 2024 Ph.D. Defense
Yuya Yamamoto

Superviser:
Nobutaka Suzuki
Hiroko Terasawa
Hiroyoshi Ito

Committee:
Nobutaka Suzuki
Hiroko Terasawa
Atsushi Toshimori
Shuichi Moritsugu
Juhan Nam