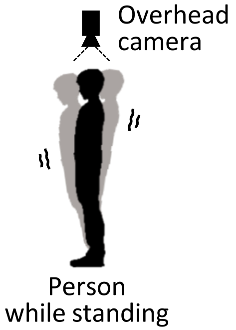
Extracting Temporal Features Robust to Headwear Variation from Video Sequences of Body Sway for Person Identification

T. Kamitani, H. Nakayama, M. Nishiyama
Tottori University, Japan



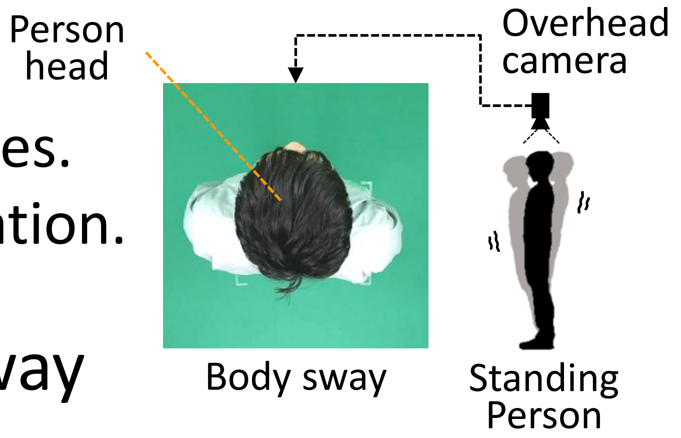
Q. Are they the same person or not?

鳥取大学
Tottori University

Introduction

Person identification

- Determines whether the same people appear in video sequences.
- Body sway has attracted attention as a cue for person identification.



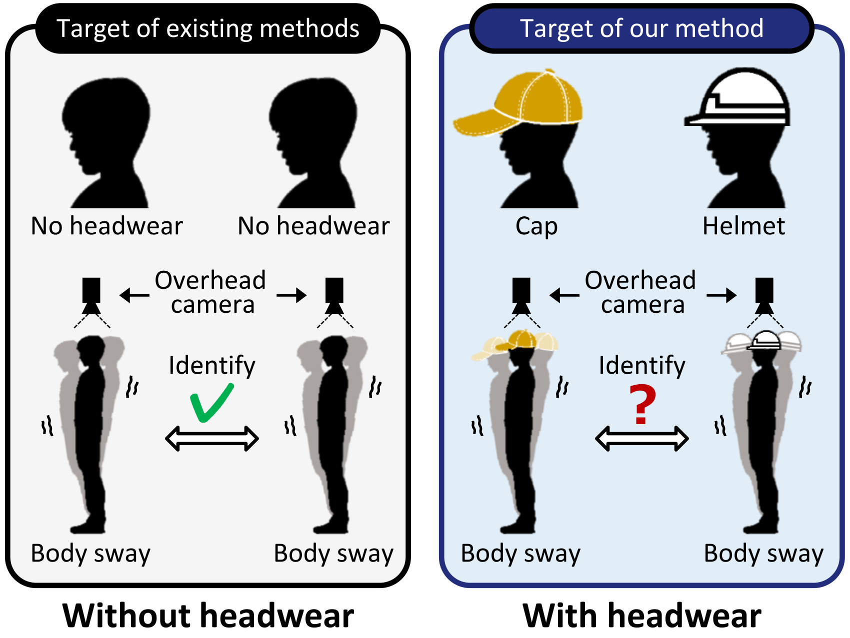
Existing methods for person identification using body sway

- Extract spatiotemporal features from overhead camera.
- require that people do not wear headwear.

People with headwear

- Factory workers may wear caps or helmets.
- Costume participants may wear wigs.

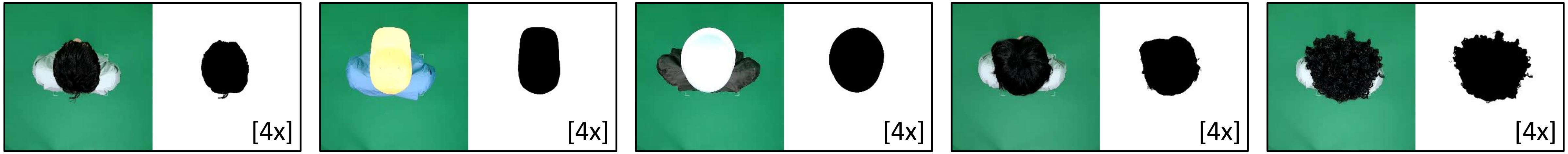
We focus on identifying people with headwear in video sequences of body sway.



When people wear headwear

People’s Head shapes change with the type of headwear worn.

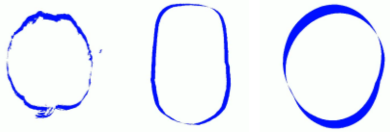
No headwear | Cap | Helmet | Long-hair wig | Afro wig



Video | Head shape | Video | Head shape | Video | Head shape | Video | Head shape | Video | Head shape

Problems with existing methods

- The spatiotemporal features capture personal identity through spatial shape and temporal movement representations.
- Because both representations rely on the head shape, the spatiotemporal features vary with changes in head shape.

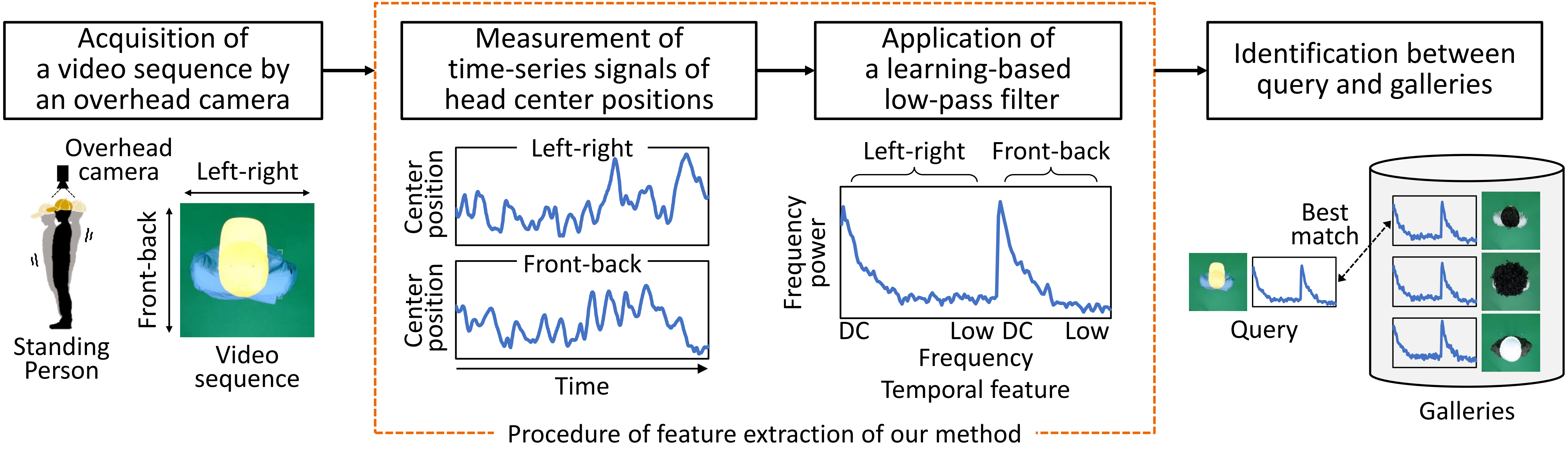


Head shapes and head movements represented by existing methods

Person identification accuracy decreases when using existing spatiotemporal features for people wearing different headwear.

Purpose

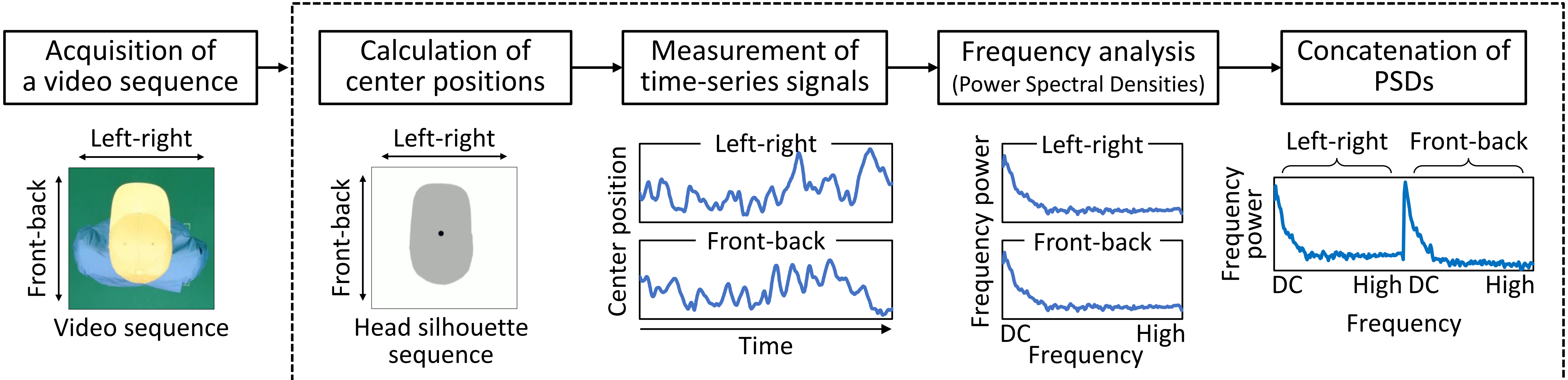
We propose a method to extract temporal features robust to headwear variation in person identification using body sway, without relying on head shape.



Our temporal features provide higher person identification accuracy than existing spatiotemporal features in the presence of headwear variation.

Design of temporal features

- We consider extracting temporal features that focus only on the head movement ignoring the head shape.
- We improve the temporal feature extraction of existing methods.



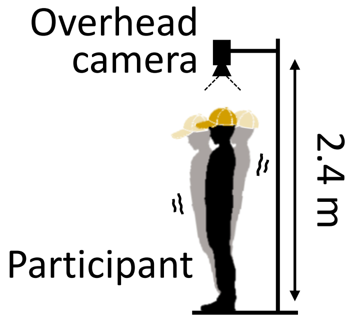
The procedure of feature extraction

- We refer to this feature extraction as **the improved existing method**.

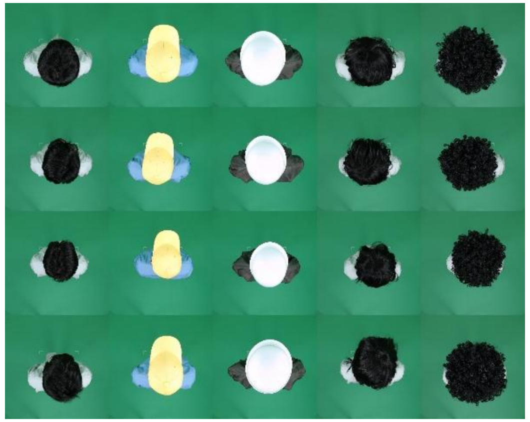
Experiment with the improved existing method

- We created an original that includes headwear variation.
  - Headwear: 5 types
  - Participants: 50
  - Observations: 2 per headwear
  - Duration: 120 sec per observation
  - Posture: Upright

No headwear | Cap | Helmet | Long hair | Afro



observation setup



Examples in the original dataset

- We evaluated person identification accuracy with headwear variation.
  - Headwear combinations in query and gallery: ${}_5P_2 = 20$ permutations
  - Classifier: Nearest neighbor algorithm
  - Evaluation metric: n-th matching rate (%)

The improved existing method slightly enhances the existing methods.

Comparison of the accuracy (%) from 50 participants

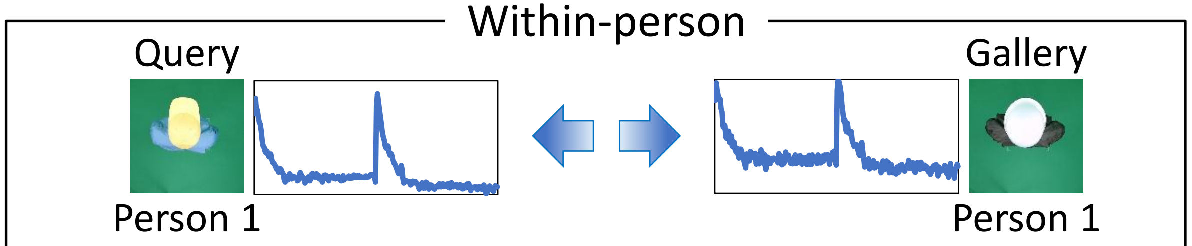
Method	**n=1**	**n=5**	**n=10**	**n=15**
Existing methods (spatiotemporal)	3.7	14.2	26.4	36.2
Improved existing method (temporal)	5.2	18.8	43.3	59.7

鳥取大学 Tottori University

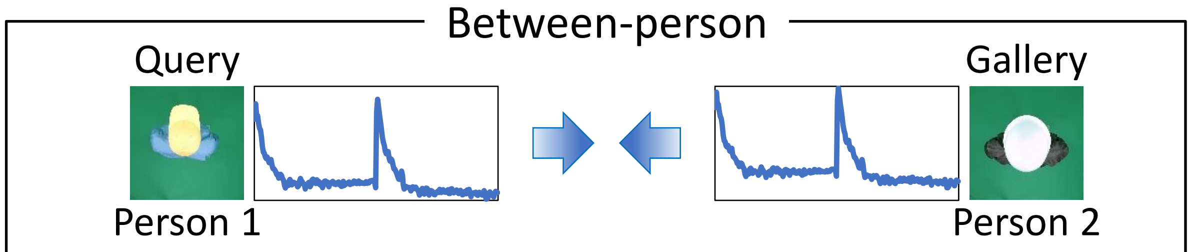
Influence of headwear variation on temporal features

We assume the improved method didn't enhance accuracy because temporal features contain uninformative components due to headwear variation.

- When uninformative components exist.



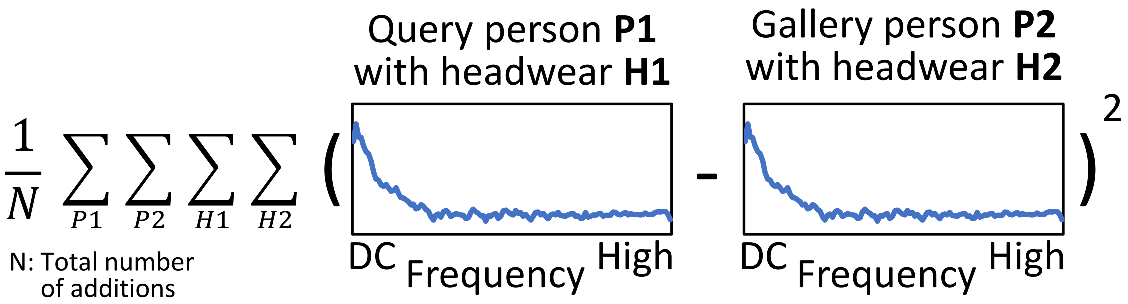
The distance between the temporal features may be large.



The distance between the temporal features may be small.

We believe these uninformative components may cause incorrect identification.

- How to investigate uninformative components in temporal features.
  - ➢ We calculated the within-person and between-person distances from the temporal features.

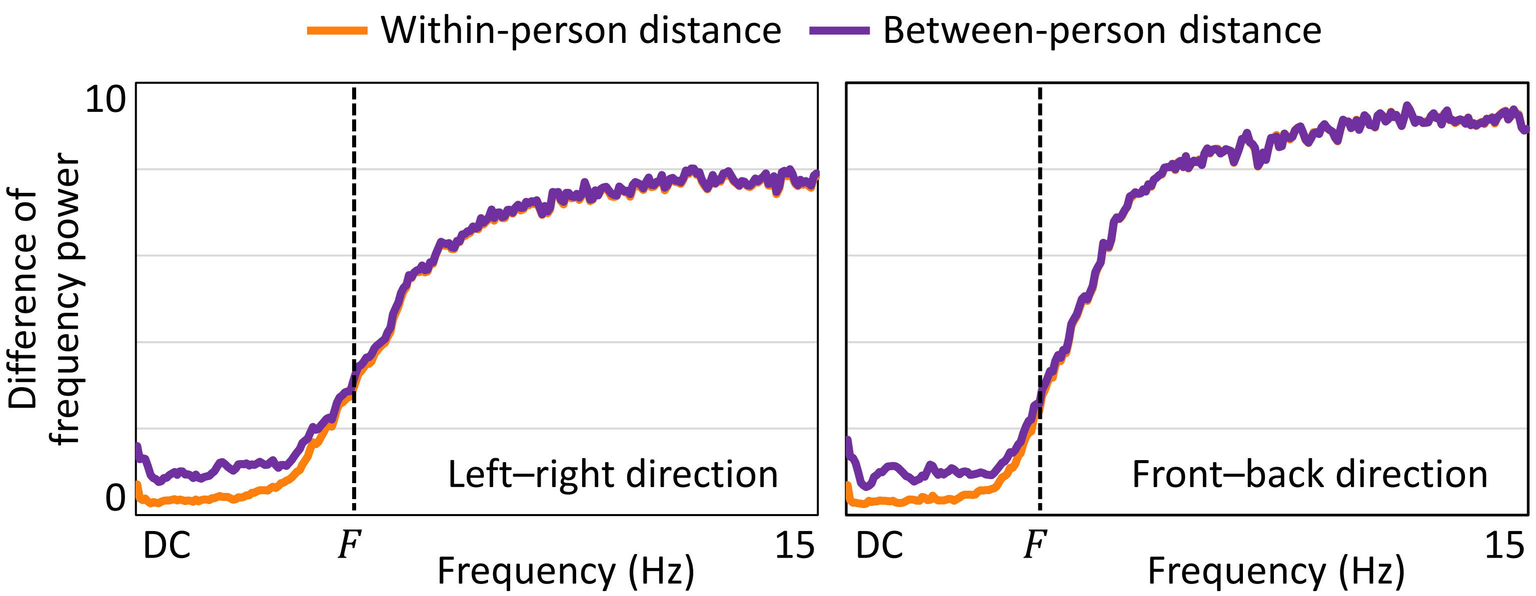


  - ➢ When P1=P2, the within-person distance.
  - ➢ When P1≠P2, the between-person distance.

We believe frequency bands where these distances are similar contain uninformative components.

Comparison of within-person and between-person distances

We compare the within-person and the between-person distance in the temporal features of the improved existing method.

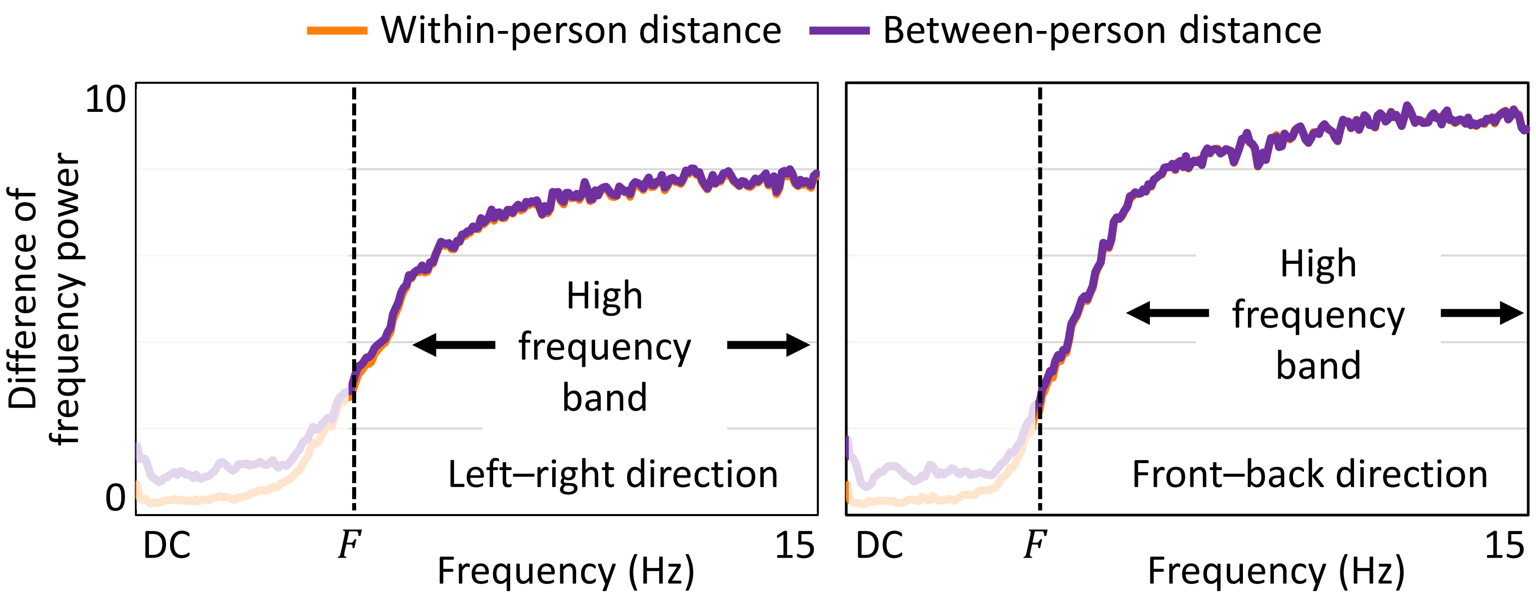


- The difference is small in the frequency range above F-Hz (High-frequency band).
- The difference is large in the frequency range below F-Hz (Low-frequency band).
- The boundary of the trend change is around F-Hz.

High-frequency components are uninformative and should be removed, while low-frequency ones are informative and should be retained.

Comparison of within-person and between-person distances

We compare the within-person and the between-person distance
in the temporal features of the improved existing method.

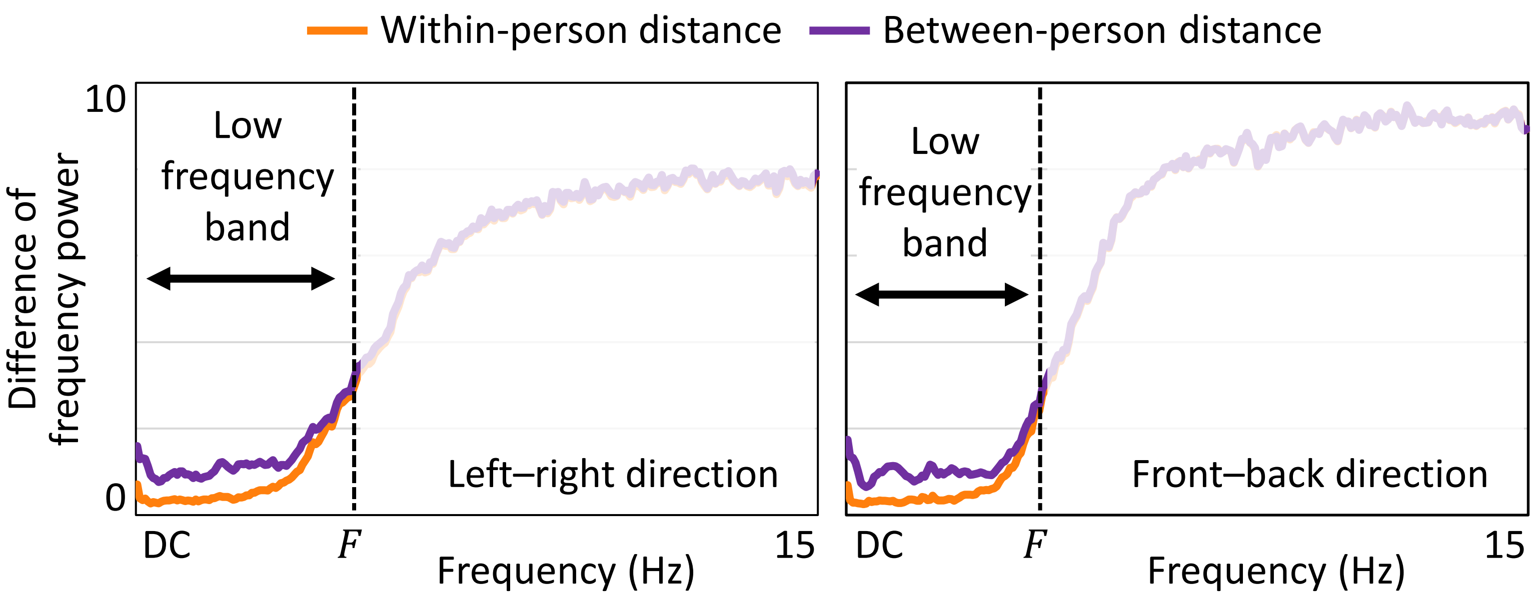


- The difference is small in the frequency range above F-Hz (High-frequency band).
- The difference is large in the frequency range below F-Hz (Low-frequency band).
- The boundary of the trend change is around F-Hz.

High-frequency components are uninformative and should be removed,
while low-frequency ones are informative and should be retained.

Comparison of within-person and between-person distances

We compare the within-person and the between-person distance
in the temporal features of the improved existing method.

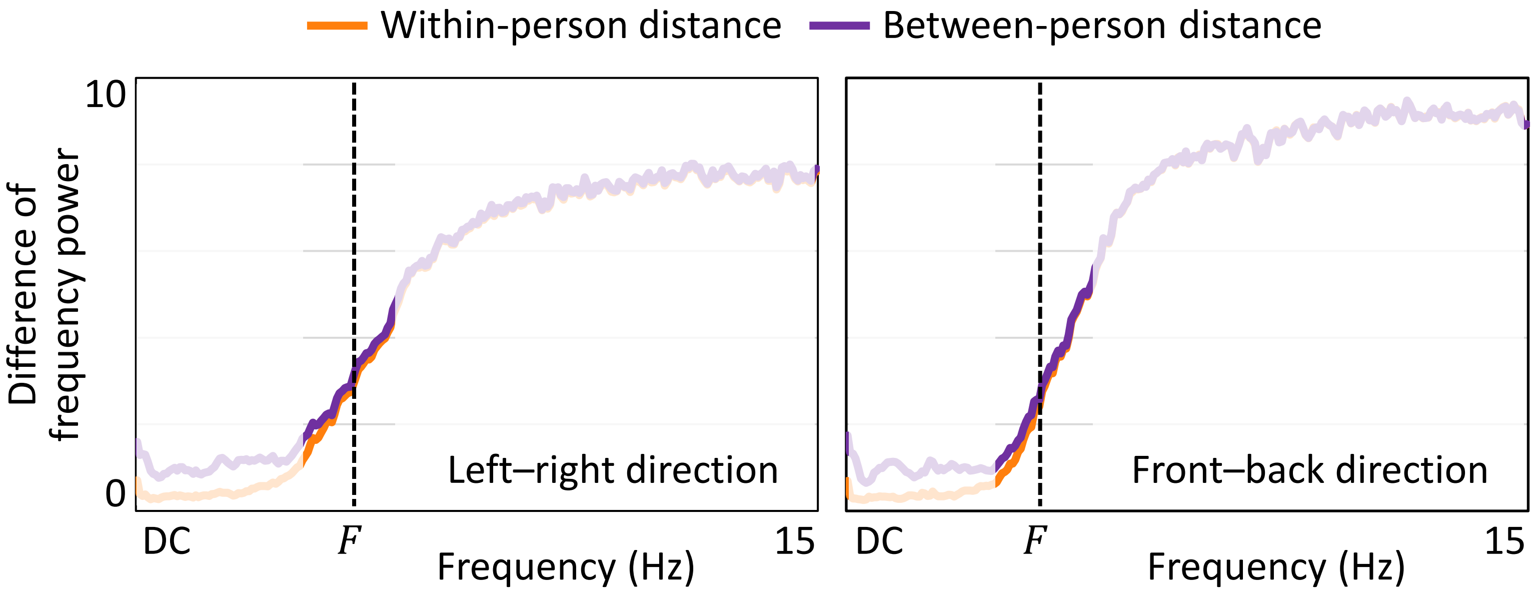


- The difference is small in the frequency range above F-Hz (High-frequency band).
- The difference is large in the frequency range below F-Hz (Low-frequency band).
- The boundary of the trend change is around F-Hz.

High-frequency components are uninformative and should be removed,
while low-frequency ones are informative and should be retained.

Comparison of within-person and between-person distances

We compare the within-person and the between-person distance
in the temporal features of the improved existing method.



- The difference is small in the frequency range above F-Hz (High-frequency band).
- The difference is large in the frequency range below F-Hz (Low-frequency band).
- The boundary of the trend change is around F-Hz.

High-frequency components are uninformative and should be removed,
while low-frequency ones are informative and should be retained.

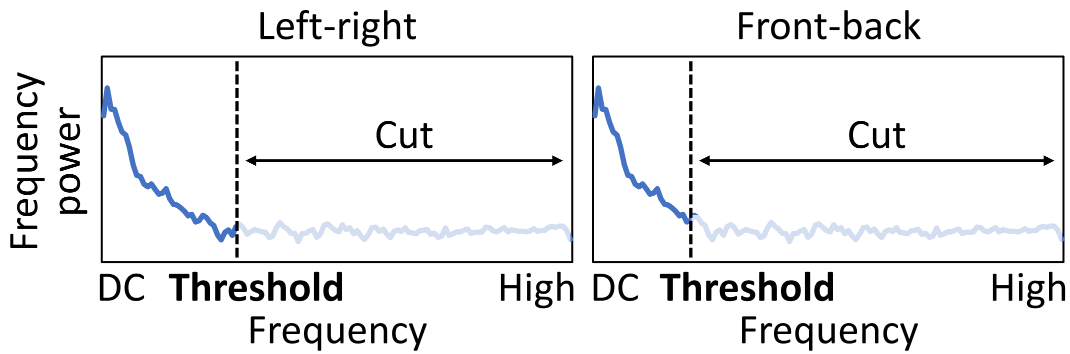
Design of a learning-based low-pass filter (1/2)

We propose a learning-based low-pass filter to extract only the informative low-frequency components for person identification.

When designing the learning-based low-pass filter

- We need to select a suitable threshold for the low-frequency band.
- We use a separation metric for selecting thresholds.

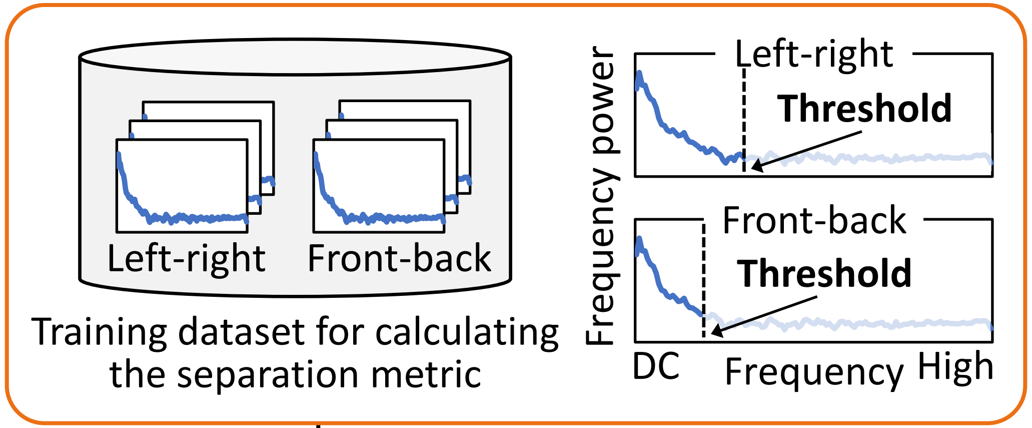
$$\text{Separation metric} = \frac{\text{Variance of between-person distances}}{\text{Variance of within-person distances}}$$

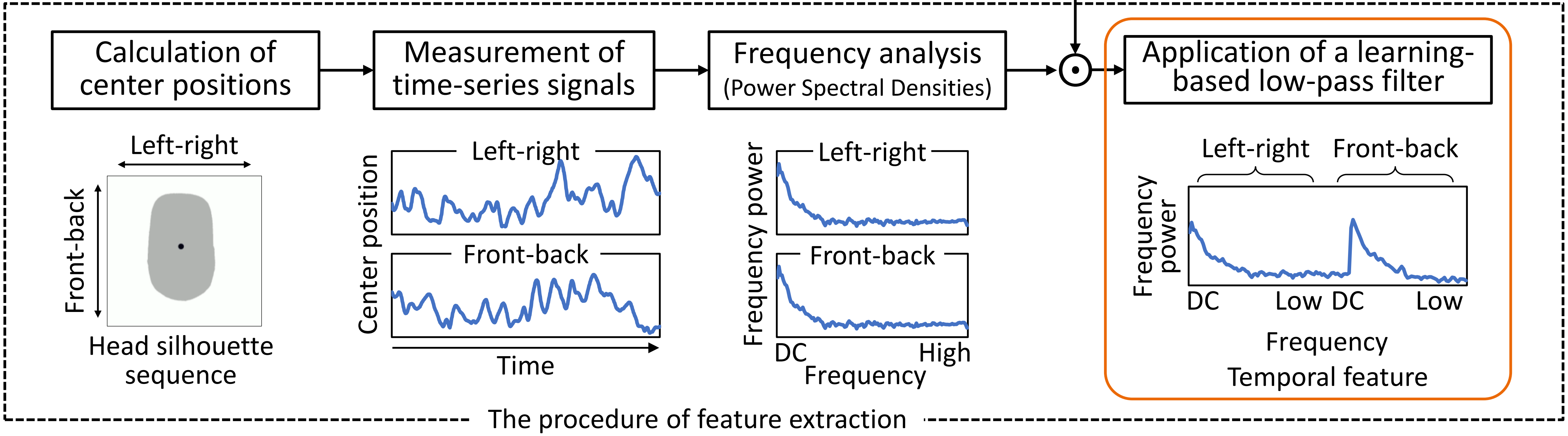


- We select the thresholds that maximize the values of the separation metric.
- The thresholds are determined separately for the left-right and front-back directions.

Design of a learning-based low-pass filter (2/2)

- To obtain robust thresholds, we prepare a training dataset.
- We refer to this feature extraction procedure as **the proposed method**.





Experiment with the proposed method

- We created training set and evaluation set from the 50 participants.
  - Training set for selecting threshold: 10 randomly selected participants
  - Evaluation set for person identification: the remaining 40 participants

- We searched for the optimal low-pass filter threshold in 0.05-Hz increments.

Comparison of the accuracy (%) from 40 participants

Method	n=1	n=5	n=10	n=15
Proposed method (with low-pass filter)	**44.2**	**78.6**	**89.2**	**94.2**
Improved existing method (without filtering)	6.3	21.9	38.1	51.0

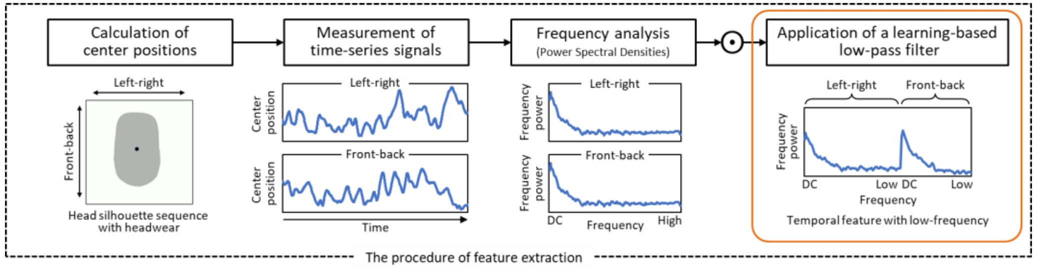
Experimental results suggest the proposed method enhances person identification accuracy with headwear variation.

Conclusions

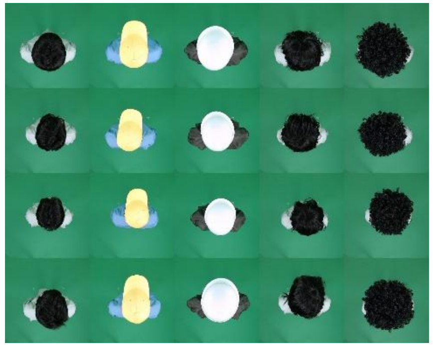
We proposed a method to extract temporal features robust to headwear variation for person identification using body sway from an overhead camera.

Our contributions

- We clarify informative and uninformative components in temporal features.



- We present a learning-based low-pass filter that removes uninformative components for robust temporal feature extraction.



- Using a dataset of people with different headwear, we demonstrate enhanced person identification accuracy with headwear variation.

In future work

We plan to identify people wearing other types of headwear.

Method	n=1	n=5	n=10	n=15
Proposed method	**44.2**	**78.6**	**89.2**	**94.2**
Existing methods	3.7	14.2	26.4	36.2