Basic Investigation for Sign Language Sentence Interpretation Using Acceleration Sensor Information

Hiroshi Tanaka[1]

Y. Umeda[1], Y. Kawakita[1], H. Nishimura[1] and J. Mitsugi[2]

[1]: Kanagawa Institute of Technology, Japan

[2]: Keio University, Japan

Contact email: h_tanaka@ic.kanagawa-it.ac.jp

Acknowledgement

This work includes the results of the research and development for expanding radio resources (JPJ000254) of the Ministry of Internal Affairs and Communications. It is also supported by "The Telecommunication Advancement Foundation", Japan.

Short resume of the presenter

Hiroshi TANAKA
Prof. Dr. Kanagawa Institute of Technology, Japan

He received Dr. Eng. at the Graduate School of Hokkaido University, Japan.
He joined Nippon Telegram and Telephone Corporation (NTT) in 1985.
He was engaged in satellite communication systems in NTT.
He has been a Professor at the Kanagawa Institute of Technology since 2006.
His current research theme is the application of IoT and AI technologies.
His team is now engaged in sign language recognition, noise removal, and sound classification by using AI technologies.

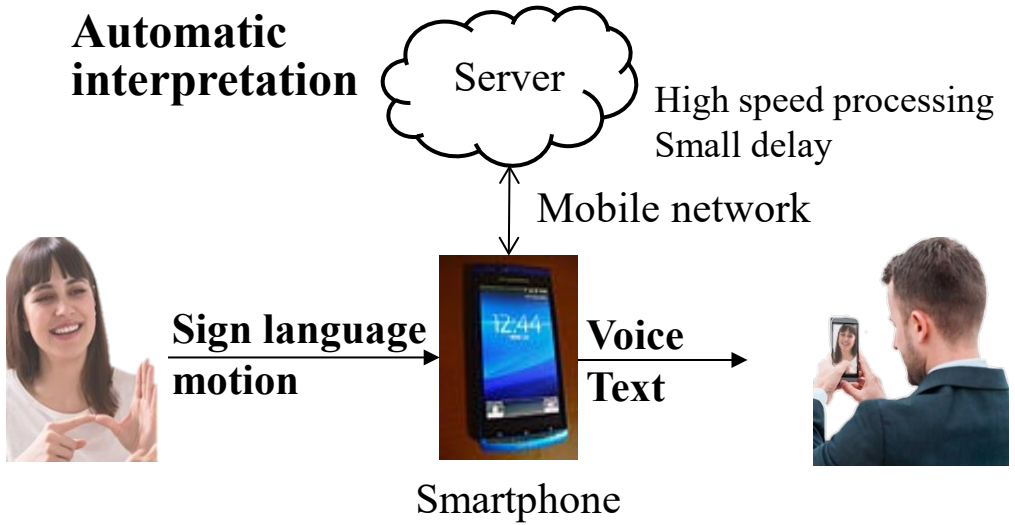
Content of today's presentation

- Research background and purpose
- Related works and features of our current approach
- Data acquisition of sign language motion
- Basic idea for segmentation of sign language sentences
- Segmentation results
- Word classification results
- Conclusion and future works

Our purpose and final goal

- Automatic interpretation of voice and text between different languages by AI
- Elimination of communication barriers between hearing-impaired and non-impaired people

Sign Language Interpretation by AI



Final stage of this research

Related work and our current approach

Research target

Classification of word motion composing sign language sentences

→ Interpretation of sentence motion [1),2)]

Big barrier

Gathering sign language sentence motion data

Gathering word motion data << sentence motion data

- Database [3),4)] of word motion of Japanese sign language

Our current approach

Sentence → Segmentation into each word

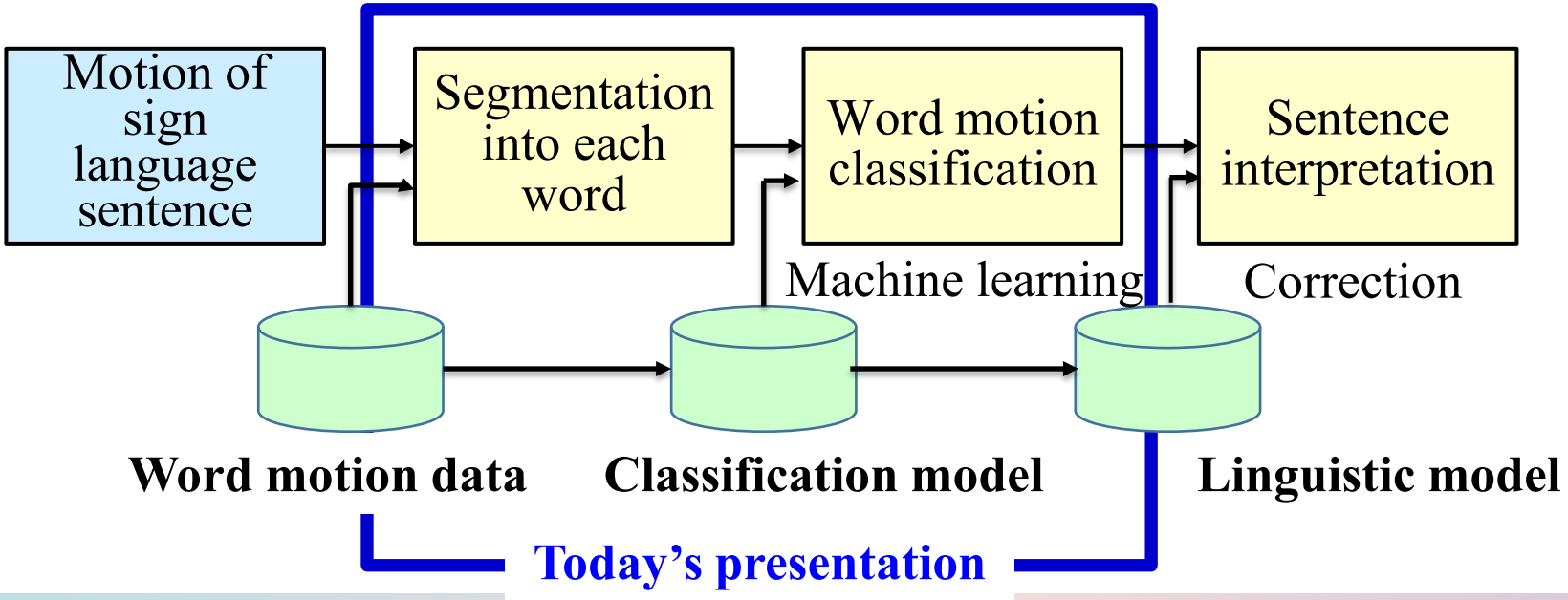
SmartDeaf[3)]

Syuwashower[4)]

[1] P. Villegas and L. Francisco, "Sign Language Segmentation Using a Transformer-based Approach.". Available from: https://hdl.handle.net/20.500.14468/14662, 21 pages, 2022.
[2] J. Huang et al., "Video-based sign language recognition without temporal segmentation." Proc. of the AAAI Conf. on Artificial Intelligence. Vol. 32. No. 1. pp. 2257-2264, 2018.
[3] SmartDeaf, https://www.smartdeaf.com/, [4] Sign Language Shower, https://www.nhk.or.jp/school/tokkatsu/syuwashower/

Our scheme and today's topic

- Segmentation of sentence motion into each word motion
- Classification (Interpretation) of words
- Correction and prediction by linguistic model



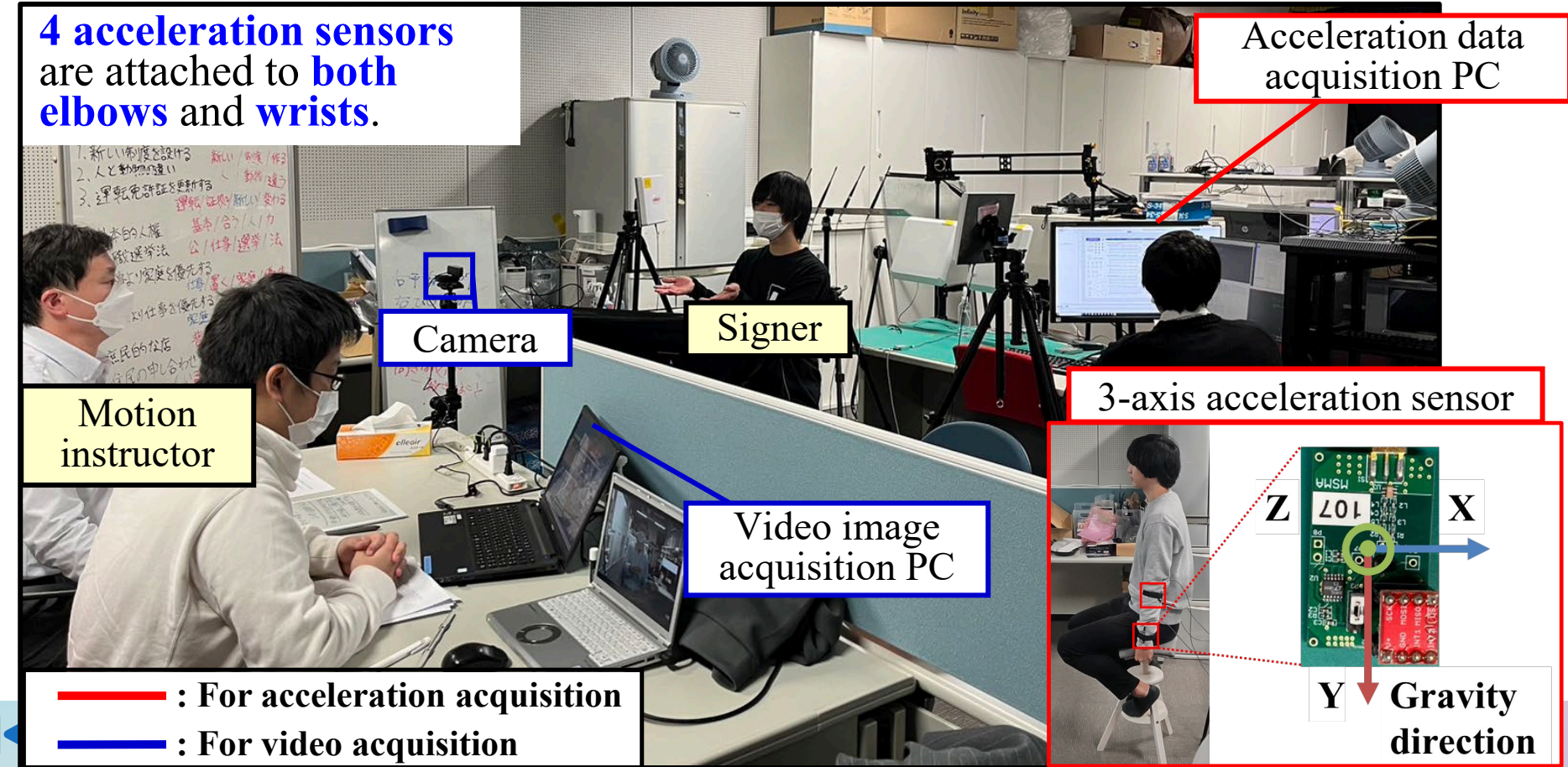
Selection of words and creation of sentences

Selection of frequently referenced words*): 22 words					
1. new	2. system	3. create	4. human	5. animal	6. difference
7. driving	8. license	9. update	10. family	11. put	12. work
13. prioritize	14. ordinary	15. people	16. familiar	17. shop	18. basic
19. power	20. public	21. election	22. law		

*) Japanese sign language dictionary

Sign language short sentences (Combination of above words): 5 sentences	
1. new/system/create (I create a new system.)	2. human/animal/difference (Humans and animals are different.)
3. driving/license/new/update (I update a driving license.)	4. family/put/work/prioritize (I prioritize work over family.)
5. ordinary/people/familiar/shop (A shop is familiar to ordinary people.)	

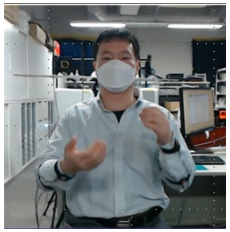
Data acquisition configuration



Example of sign language sentence motion

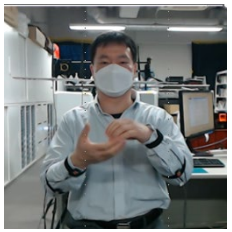
Sentence motions are composed of words and transition motions.

Driving

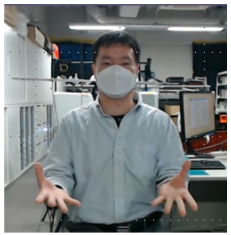


Transition

License



Transition

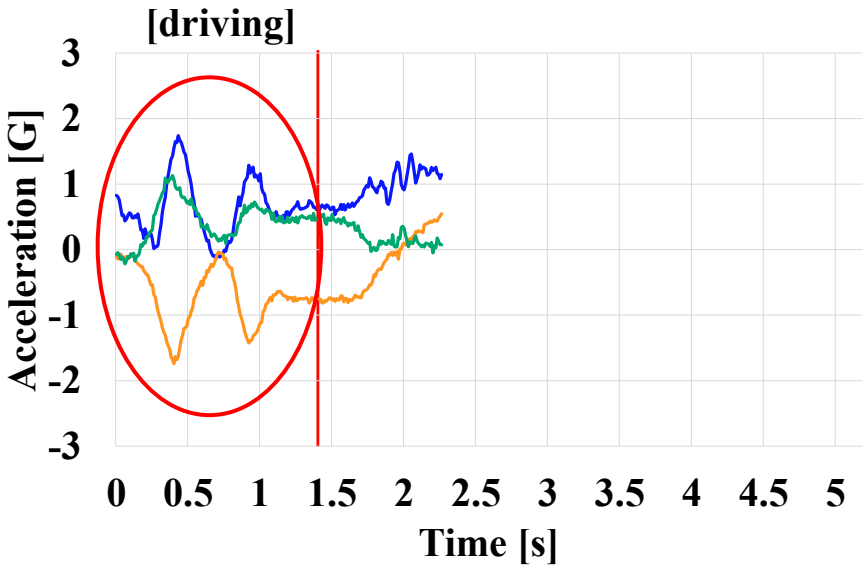


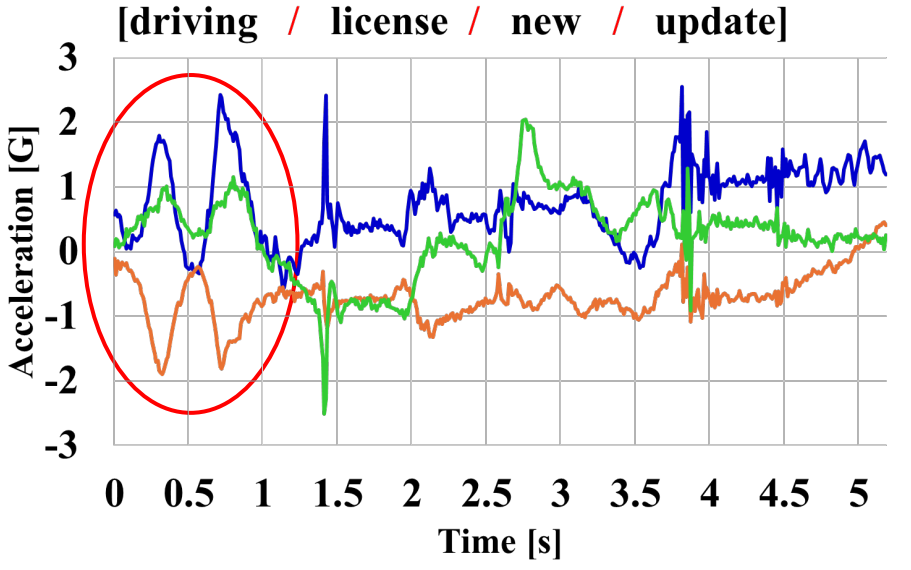
New

Transition

Update

Acceleration data waveform and segmentation





Left wrist X-axis　Left wrist Y-axis　Left wrist Z-axis

Sentence motion is composed of word motion including transition motion.
→The region whose shape is similar to the word motion is included.
Method: Likelihood by **LSTM model** or Distance by DTW (DP matching)

Likelihood from LSTM model

Input
Sentence motion data
(Acceleration data)

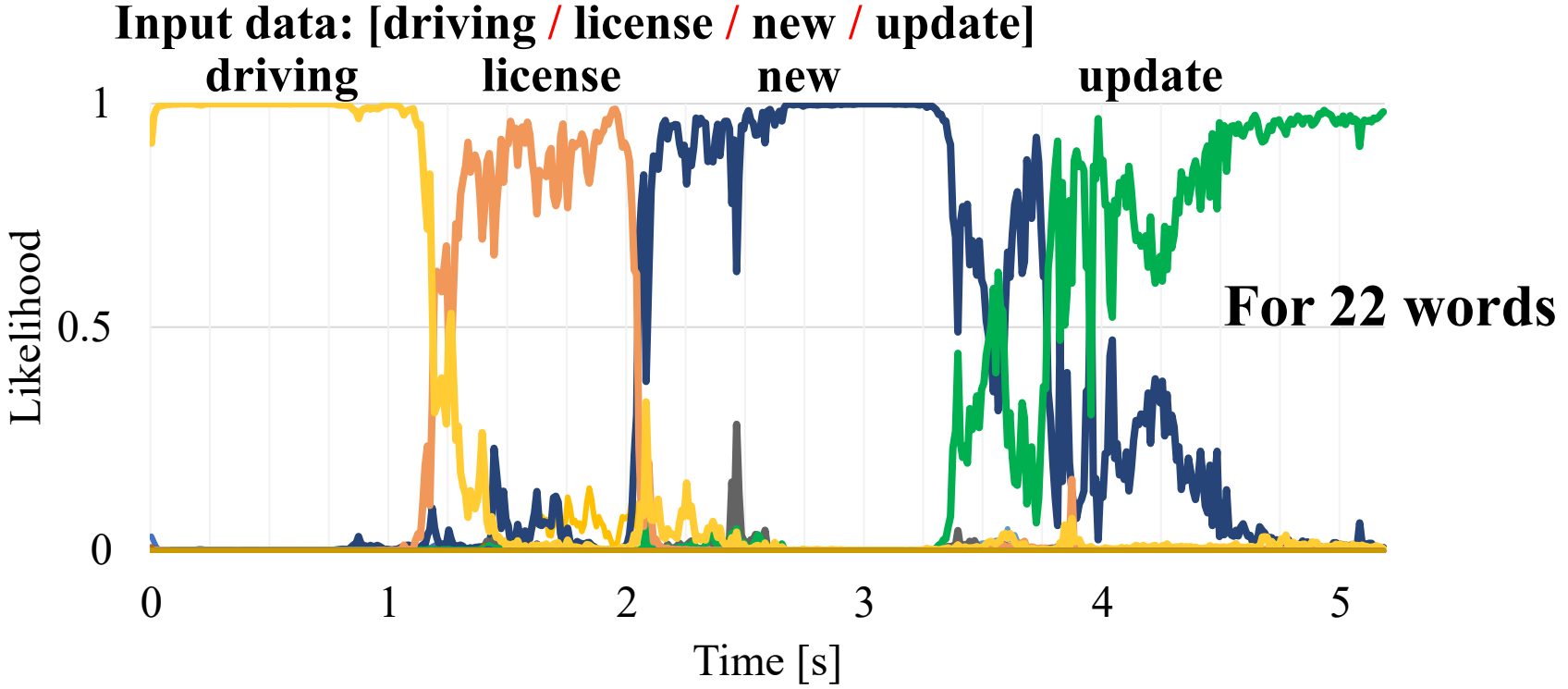
→ **LSTM model by word motion data** →

Output
Likelihood of each element of word motion

Signer	Num of words	Num of samples/ word	Total samples
1	22	15	330

Sign language short sentences (Combination of above words): 5 sentences, 3 samples each	
1. new/system/create (Create a new system.)	2. human/animal/difference (Humans and animals are different.)
3. driving/license/new/update (Update a driving license.)	4. family/put/work/prioritize (Prioritize work over family.)
5. ordinary/people/familiar/shop (A shop is familiar to ordinary people.)	

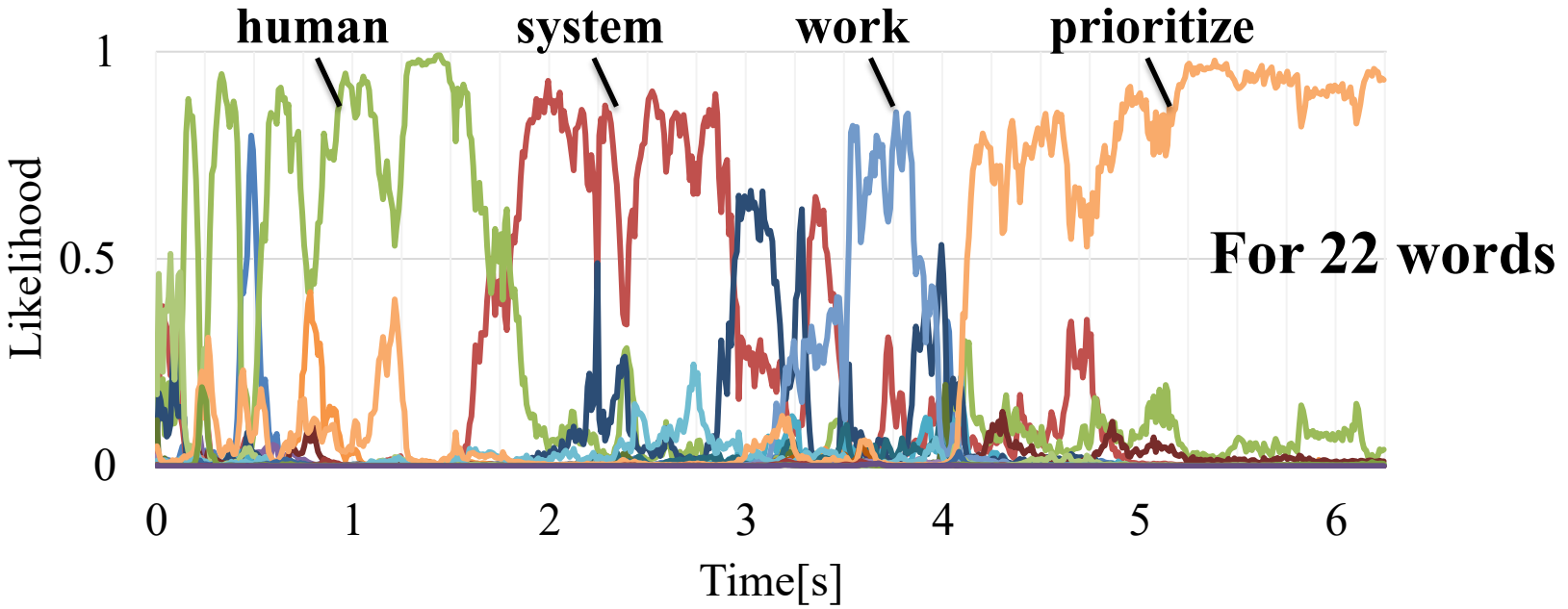
Example of Likelihood from LSTM model (1)



Likelihood changes due to word differences are clear, and segmentation is easy.

Example of Likelihood from LSTM model (2)

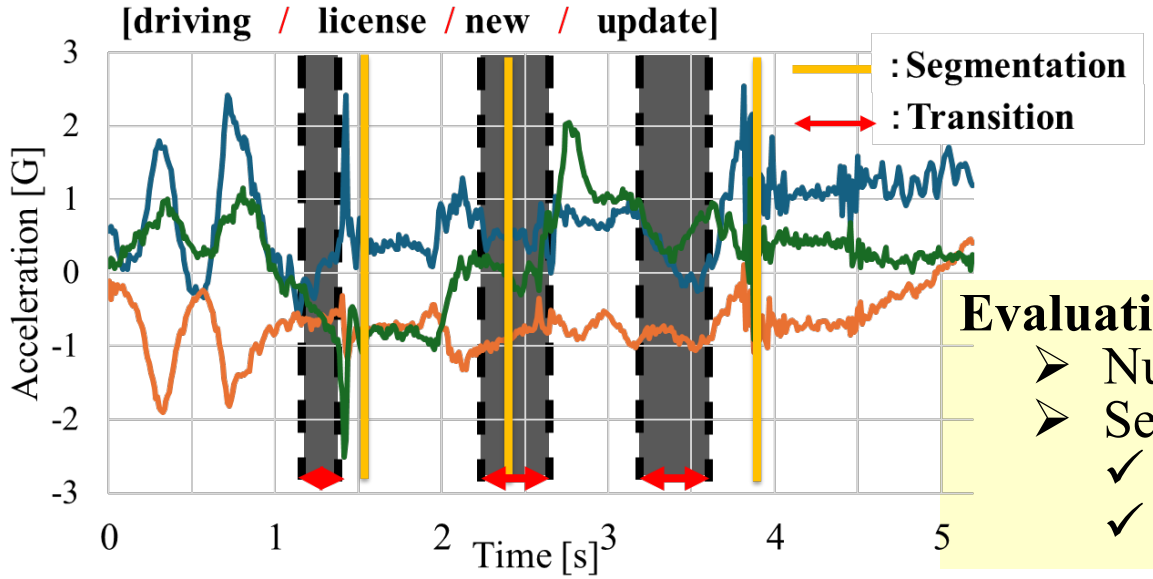
Input data: [family / put / work / prioritize]



The change in likelihood is unstable and segmentation is not easy.
The likelihood of some words is higher than that of actual words.

Segmentation method and example of segmentation result

Segmentation method:
1) Accumulation of the likelihood of each word
2) Segmentation at the point where the increase of the highest value word is saturated
3) Neglecting the length of the segmentation span < 0.3 sec



Evaluation

- Number of divisions: 3 (Correct)
- Segmentation position
  - ✓ Correct: 1/3
  - ✓ Incorrect: 2/3

Evaluation result for segmentation

Short sentences	Num of words	Num of Segment			Segment position index*		
1. new/system/create	3	4	3	3	0.25	0.33	0.67
2. human/animal/difference	3	3	3	4	0.00	0.33	0.25
3. driving/license/new/update	4	4	4	4	0.25	0.50	0.50
4. family/put/work/prioritize	4	4	3	7	0.00	0.33	0.43
5. ordinary/people/familiar/shop	4	4	5	5	0.25	0.20	0.40

(**three sample sentences** for each short sentence)

*) Segment position index:

Number of segmentation positions in the transition section / number of segments

The ratio of correctly split is 60%.

The ratio of correct split position is less than 50%.

Word classification results based on segmentation

Classification model: SVM
Segmented data is classified by SVM model.
SVM models are created with 22 kinds of word data, a total of 330 data.
Motion data: 10 division, Feature elements: Average, Standard dev.

Example of classification result （driving / license / new / update)

Sentence	1st seg.	2nd seg.	3rd seg.	4th seg.
driving/license/new/update	driving	familiar	new	system
	people	license	people	update
	law	law	driving	driving

The top three classification results for 22 words are shown in this table.
All words are selected in the top 3.

Word classification results for segmented sections

Evaluation Index (*EI*) = α / β

α : number of words correctly classified, β : number of segments

Short sentences	First place only			Up to 3rd place		
1. new/system/create	0.50	0.33	0.66	0.50	0.66	1.00
2. human/animal/difference	0.33	0.66	0.25	0.33	0.66	0.75
3. driving/license/new/update	0.50	0.75	0.75	1.00	1.00	1.00
4. family/put/work/prioritize	0.50	0.66	0.28	1.00	1.00	0.42
5. ordinary/people/familiar/shop	0.25	0.80	0.40	0.75	0.80	0.60
Average (Each sentence has 3 samples)	0.51			0.76		

> 0.5

The contained words in sentence can be classified.
The possibility of the proposed scheme for interpretation seems to be shown.

Conclusion and future works

Propose of architecture of interpretation of sign language sentences

- Segmentation based on similarity of word and sentence motion
- Word motion classification based on segmentation results

Feasibility is considered to be confirmed, although only 22 kinds of words and 5 sentences.

We will collect linguistic information, such as word frequency and word order of a signed language restricted to a specific field, and create a linguistic model for the correction of classification errors.