


A METHOD FOR ESTIMATING THE GROUPING OF PARTICIPANTS IN CLASSROOM GROUP WORK USING ONLY AUDIO INFORMATION

Osamu Ichikawa (Shiga University, Japan), Takahiro Nakayama (The University of Tokyo, Japan),
 Yuuto Shima (Shiga University, Japan), Hajime Shirouzu (National Institute for Educational Policy Research, Japan)

Background

- Monitoring group-work (active learning) in schools.
- Each student wears a close-talk microphone, to record his/her speech individually.
- Speech will be transcribed by ASR.
- Teacher will check the transcription of each group.

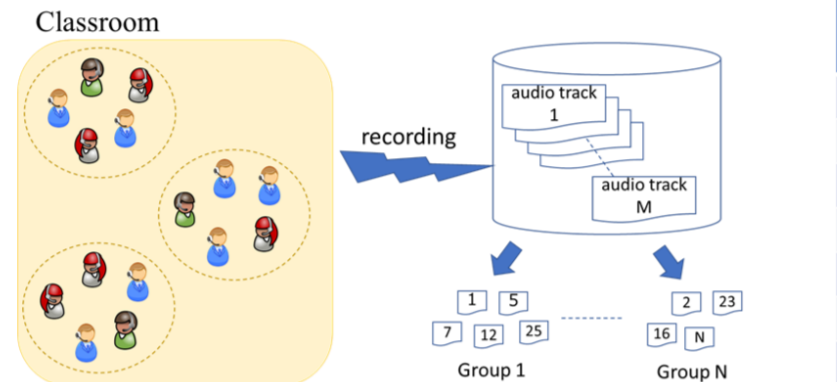


Student 1	Student 2	Student 3
Do you see ?		Yeah, perhaps.
	This is pretty.	What's up ?
This is F.		You mean about C ?
F-dash.		dash ?
		Not garbage, haha.

Problem to solve

- There are several groups in a classroom.
- There are so many recording tracks associated with each student.

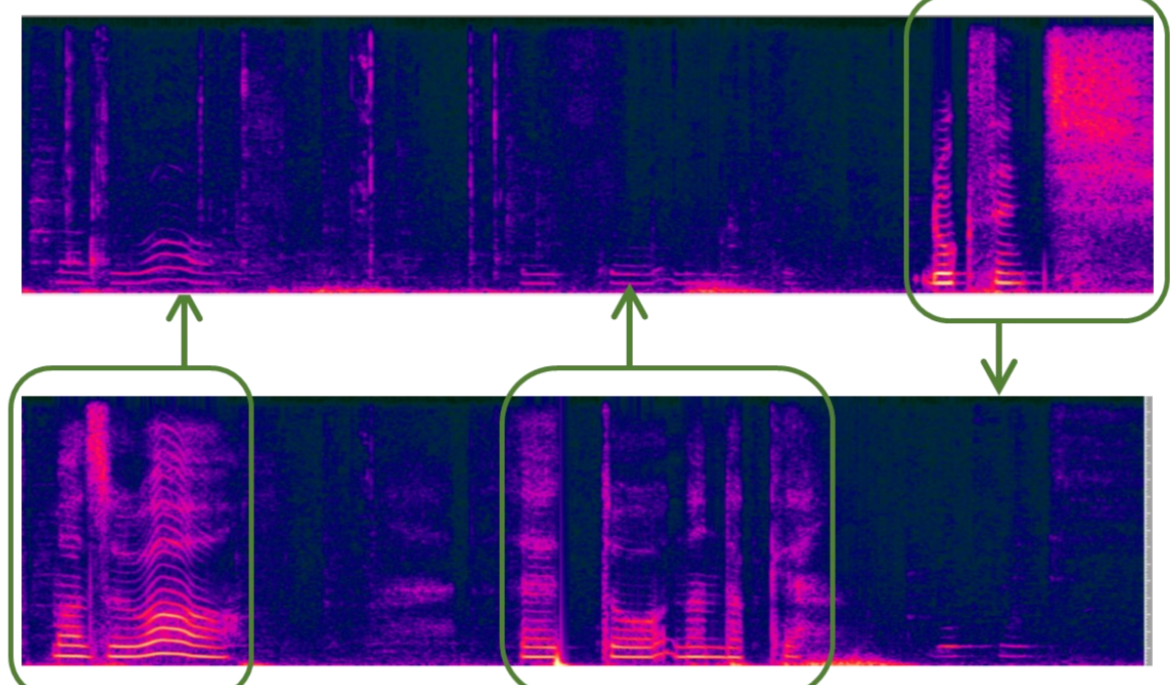
→ Teachers must maintain a table of relationships between recording tracks and groups.
 → Objective : Automatic group estimation by audio information.



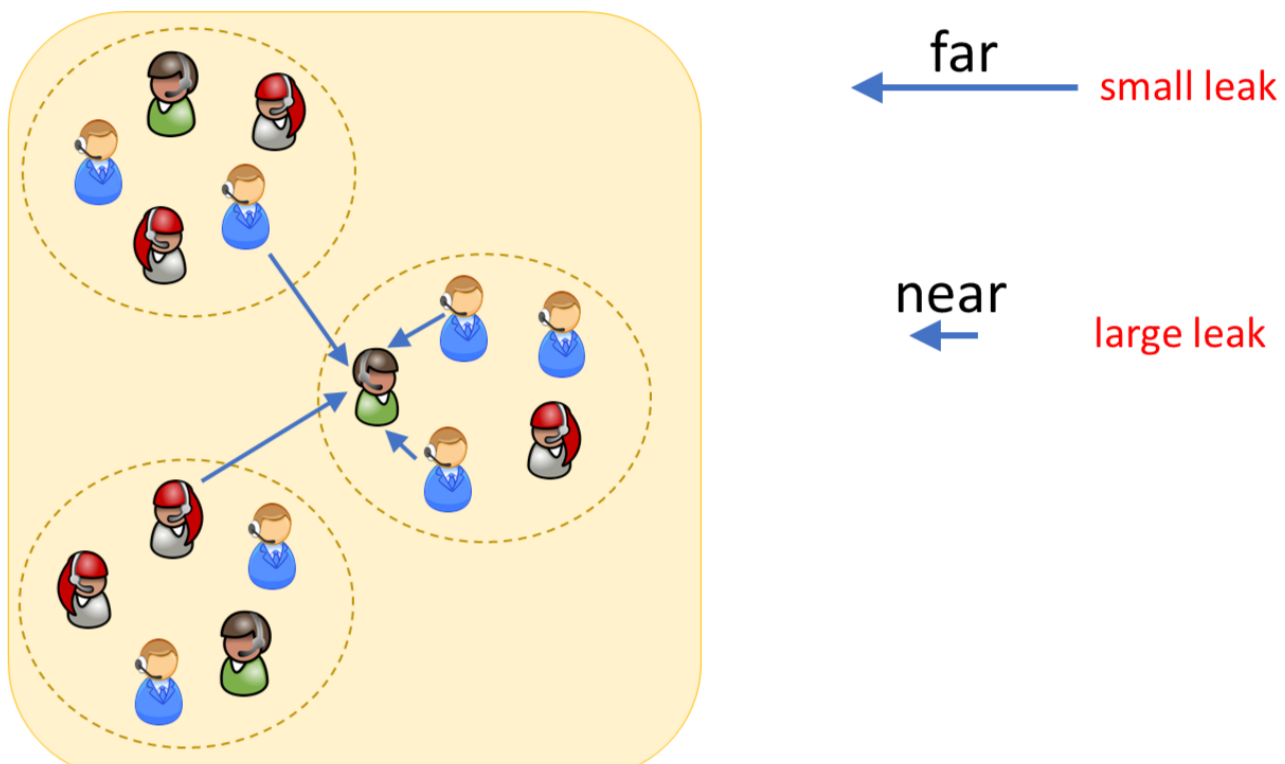
Voice Recorder ID (= Track ID)	Group ID	Voice Recorder ID (= Track ID)	Students ID
1	1	1	1
2	1	2	2
3	1	3	3
4	2	4	4
5	2	5	5
:	:	:	:
M	N	M	M

Insights for Solution

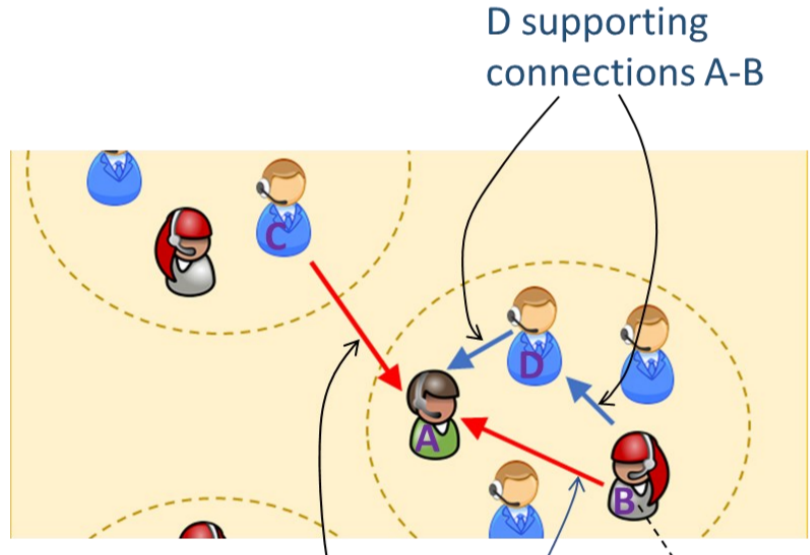
- Microphone also receives audio leaked from others



Proximity can be measured by leakage

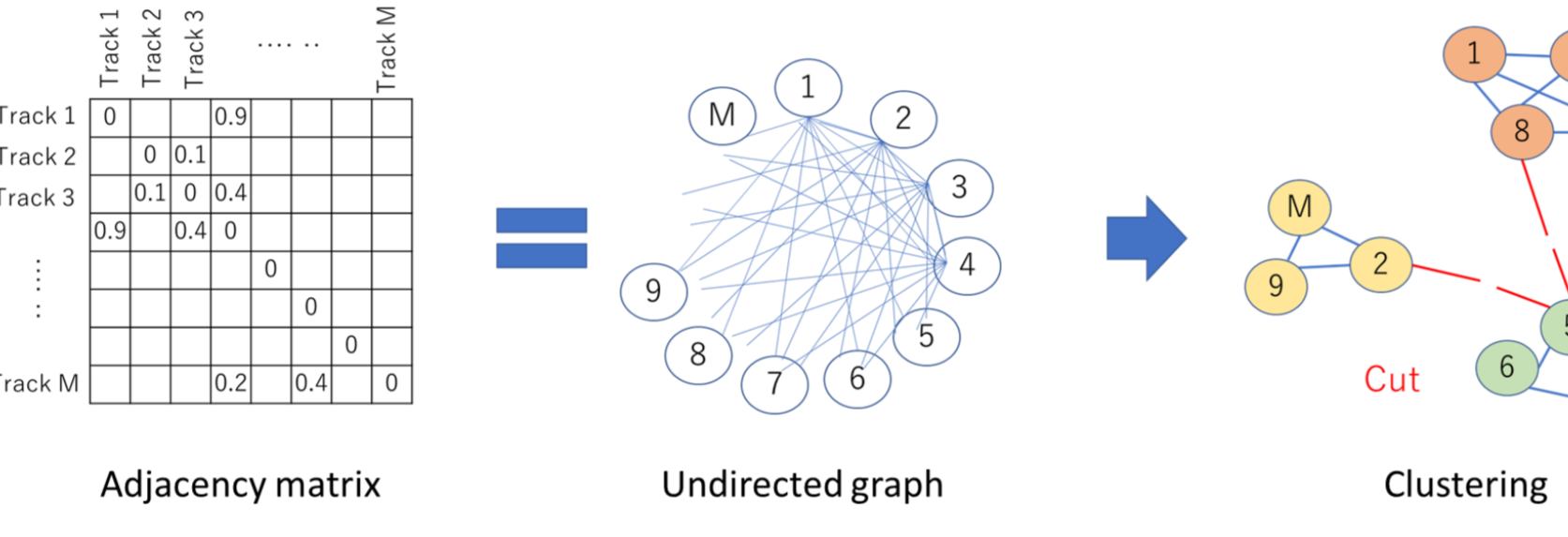


Proximity can be clustered



A-C and A-B same proximity
 A should be in the same group of B, rather than C

Network clustering approach



In this study, target number of clusters N is given.

CSP (correlation) for the metric of leakage

- Φ : cross-spectrum phase analysis (CSP)
- between track p and track q
- for t-th frame with r frames delay in track q
- using complex spectrum X

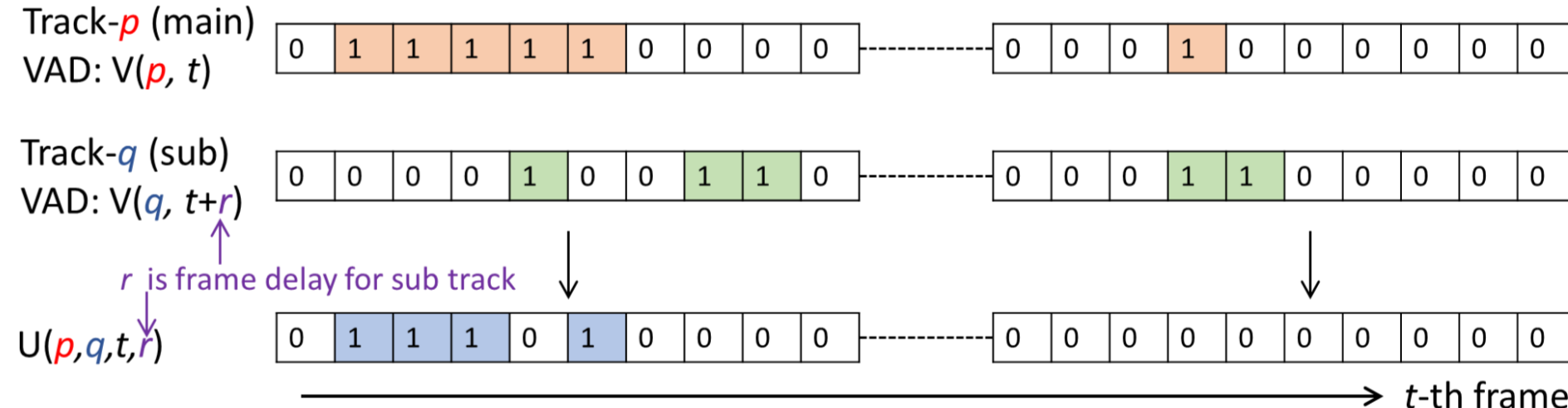
$$\phi(p, q, t, r) = IDFT \left[\frac{X(p, k, t)X(q, k, t+r)^*}{|X(p, k, t)||X(q, k, t+r)|} \right]$$

- Φ is obtained as time-domain coefficients. Take max for the time-delay d.

$$C(p, q, t, r) = \max_d \phi_d(p, q, t, r) \rightarrow \text{Integrate this in time (frame) direction}$$

Determine leak detection target frames

- Select two tracks (p and q) from all combinations. Assume recording delay r.
- The one with more active frames is the main track and the one with fewer active frames is the sub-track.



Leak detection target frames : $U(p, q, t, r) = V(p, t)(1 - V(q, t+r))$

Integrating CSP coefficients in time direction

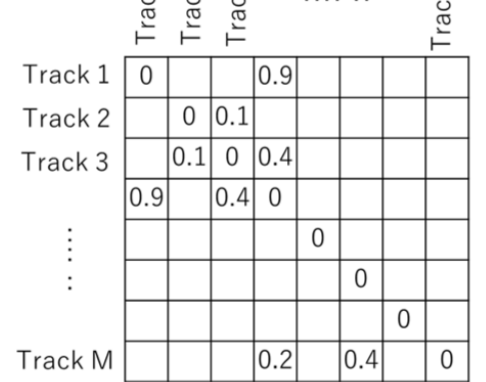
$$\hat{C}(p, q, r) = \frac{\sum_t C(p, q, t, r)U(p, q, t, r)}{\sum_t U(p, q, t, r)}$$

- Find the maximum value C_{max} by varying r.

$$C_{max}(p, q) = \max_r \hat{C}(p, q, r)$$

to adjacency matrix

Filling adjacency matrix with CSP max values

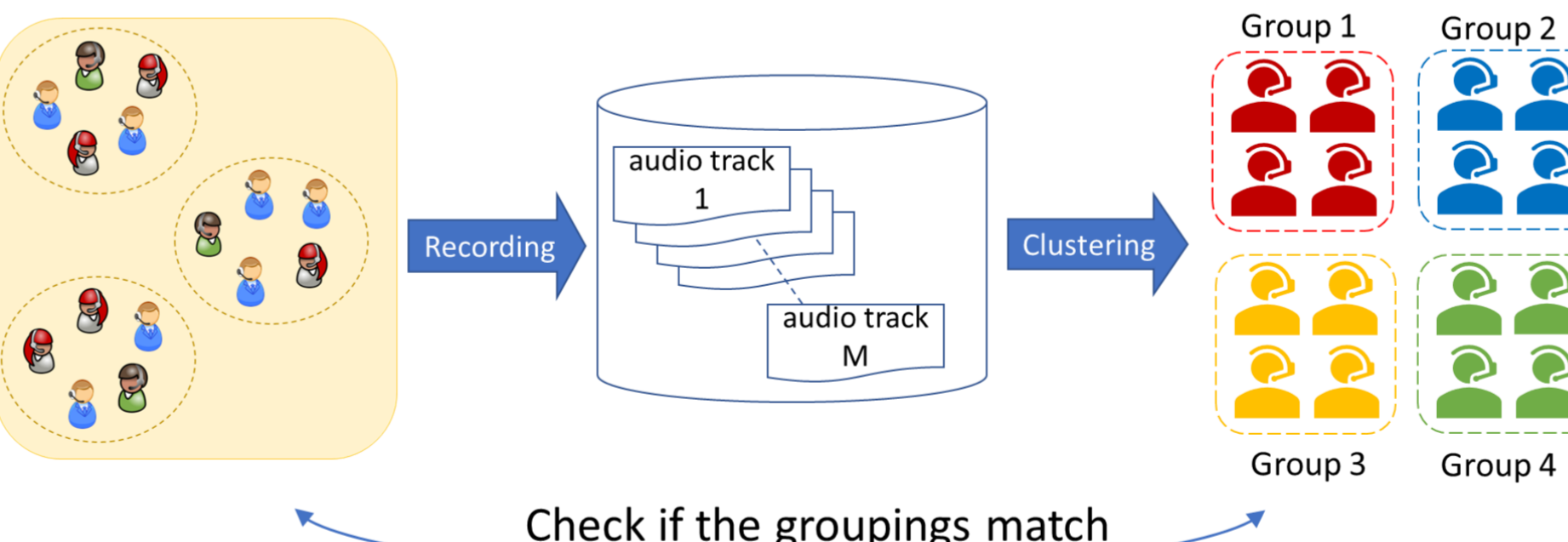


$$\alpha(p, q) = \alpha(q, p) = \begin{cases} C_{max}(p, q) & , p \neq q \\ 0 & , p = q \end{cases}$$

Spectral clustering method is performed with given target number of clusters N.

Evaluations

- We used real audio data recorded in group work classes in the actual junior high school.



Check if the groupings match

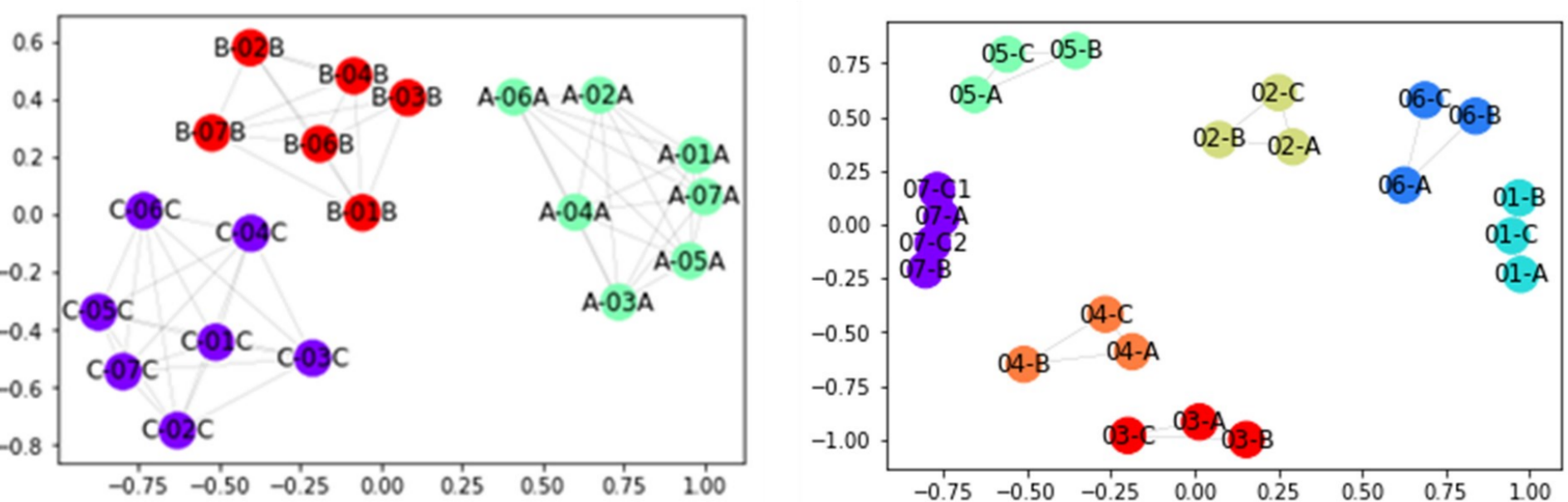
Experimental results

- We succeeded in clustering perfectly correctly in 5 out of 6 cases.

Class	Session name	Number of groups	Session time (m:s)	Result: Rand index
Mathematics	Expert	7	8:58	1.00
	Jigsaw	7	23:37	1.00
Science	Expert	3	8:35	1.00
	Jigsaw	6	20:05	1.00
Japanese	Expert	6	5:37	0.97
	Jigsaw	6	22:50	1.00

- There was only one error, one node belonged to the wrong group.

Example of clustering results



"Expert Session" of the "science" class.
 "Jigsaw Session" of the "mathematics" class.

Conclusion

- Assuming that there are multiple discussion groups in a classroom, we have proposed a novel method to estimate the grouping using only audio information, recorded by the microphones attached to each student.
- An evaluation experiment was conducted using audio data recorded in actual junior high school classes, and an average Rand index of 0.995 was obtained.
- This means that out of the six sessions, only one person was misassigned. It was verified that the method has a practical level of accuracy.