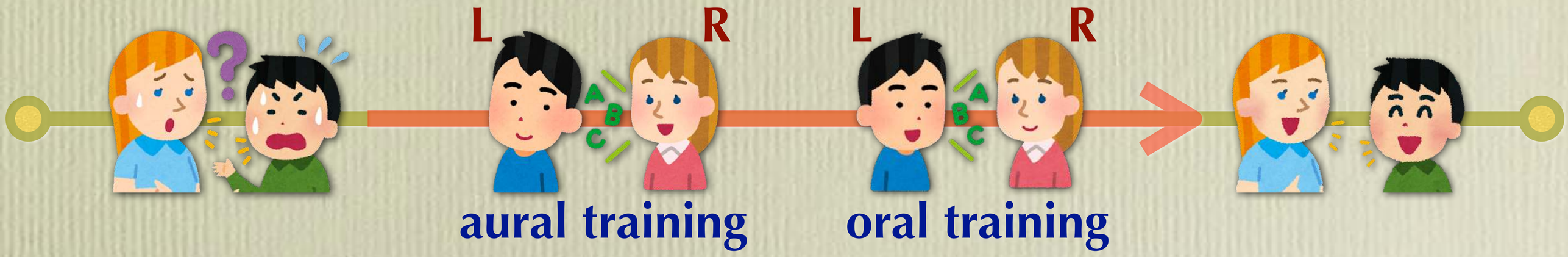
Measuring, Analyzing, and Predicting Listening Disfluency of Learners and Raters: Using Speech and AI Technologies for Automated Assessment



aural training oral training

Nobuaki MINEMATSU
Prof. of Engineering and Language Education,
The University of Tokyo

Why listening disfluency (LD)?

Changing principles in L2 pronunciation teaching

- Native-likeness
  - Degree of phonetic and prosodic similarity of learners' pronunciation to native pronunciation
  - **Speaking behaviors** are measured and analyzed with speech technology.
- Intelligibility and comprehensibility [Derwing & Munro'97]
  - Degree of listening and comprehension effort required to understand learners' speech
  - **Listening behaviors (of raters)** should be measured and analyzed, but with brain technology?

Perception or production, which should come first?

- Perceptual skills should be strengthened before production skills [Krashen'82, Shirai'08].
  - **Listening behaviors (of learners)** should be measured and analyzed, but with brain technology?





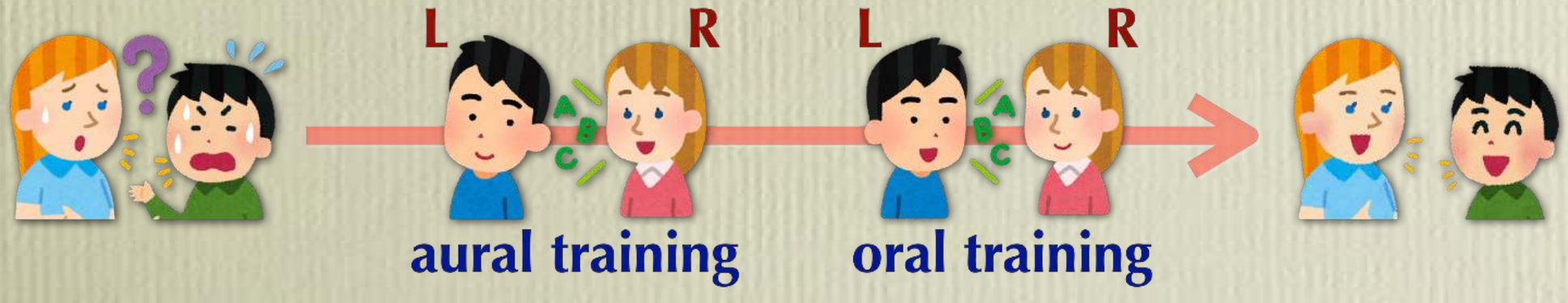
Why listening disfluency (LD)?

Changing principles in L2 pronunciation teaching

- Native-likeness
  - Degree of phonetic and prosodic similarity of learners' pronunciation to native pronunciation
  - **Speaking behaviors** are measured and analyzed with speech technology.
- Intelligibility and comprehensibility [Derwing & Munro'97]
  - Degree of listening and comprehension effort required to understand learners' speech
  - **Listening behaviors (of raters)** should be measured and analyzed, but with brain technology?

Perception or production, which should come first?

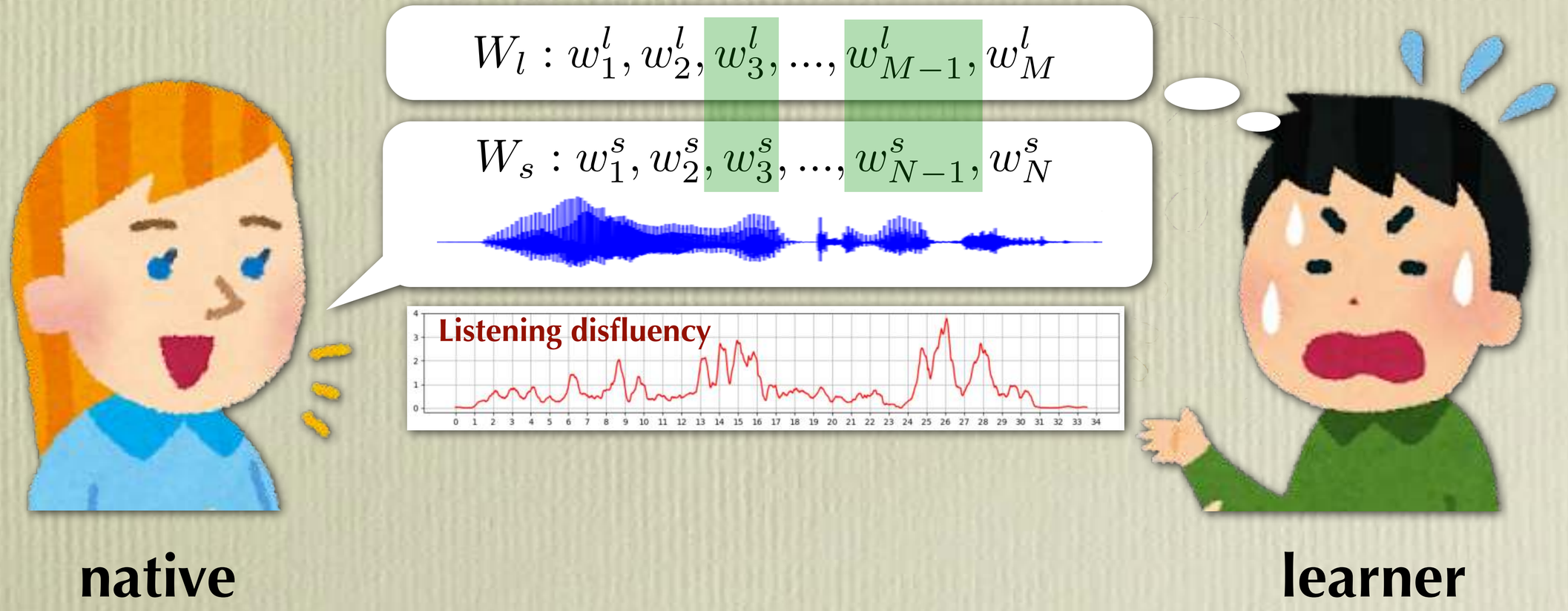
- Perceptual skills should be strengthened before production skills [Krashen'82, Shirai'08].
  - **Listening behaviors (of learners)** should be measured and analyzed, but with brain technology?



Assessment of listening [Inoue+'18]

When listening, where in a given speech does LD take place?

- Listening is mental activity and does not present any acoustic events.

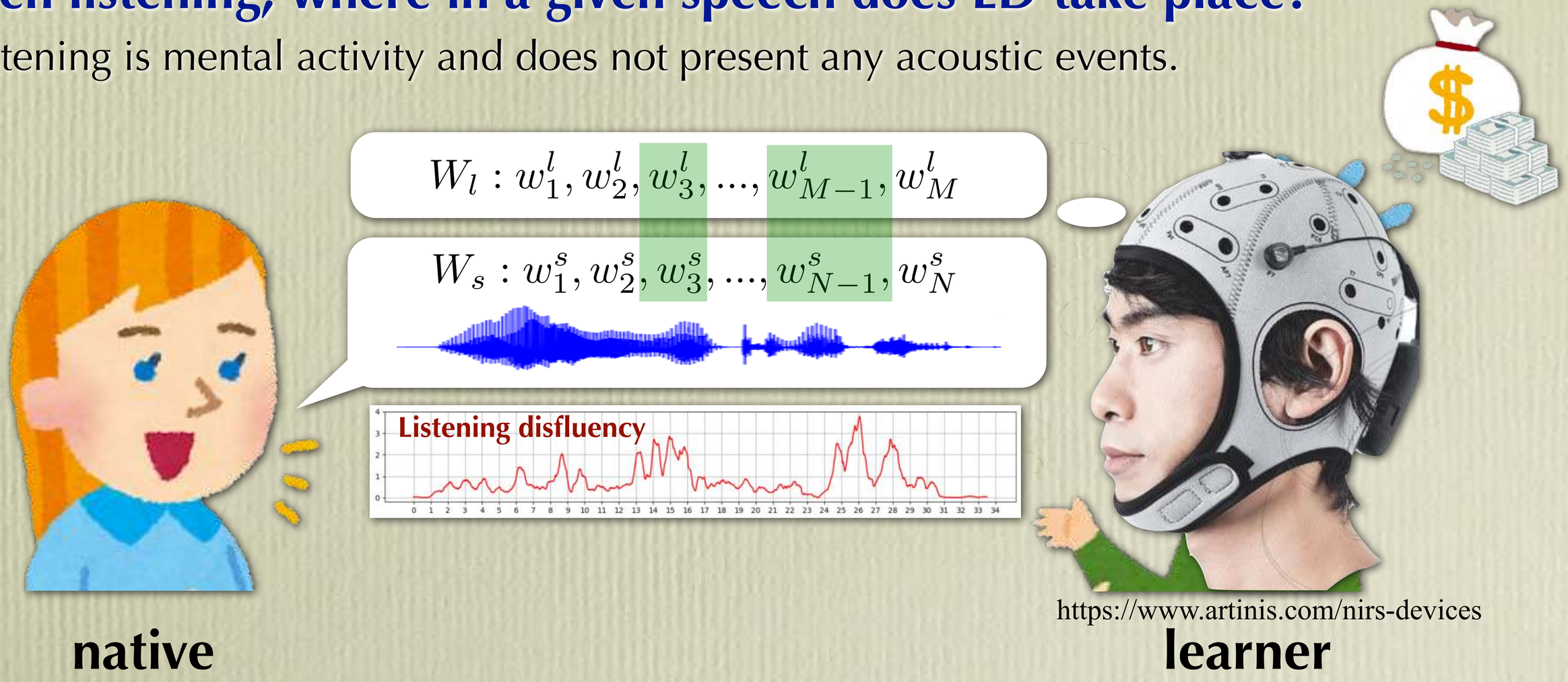


Inoue, Y., et al. (2018) A Study of Objective Measurement of Comprehensibility through Native Speakers' Shadowing of Learners' Utterances. Proc. Interspeech 2018, 1651-1655

Assessment of listening [Inoue+'18]

When listening, where in a given speech does LD take place?

- Listening is mental activity and does not present any acoustic events.

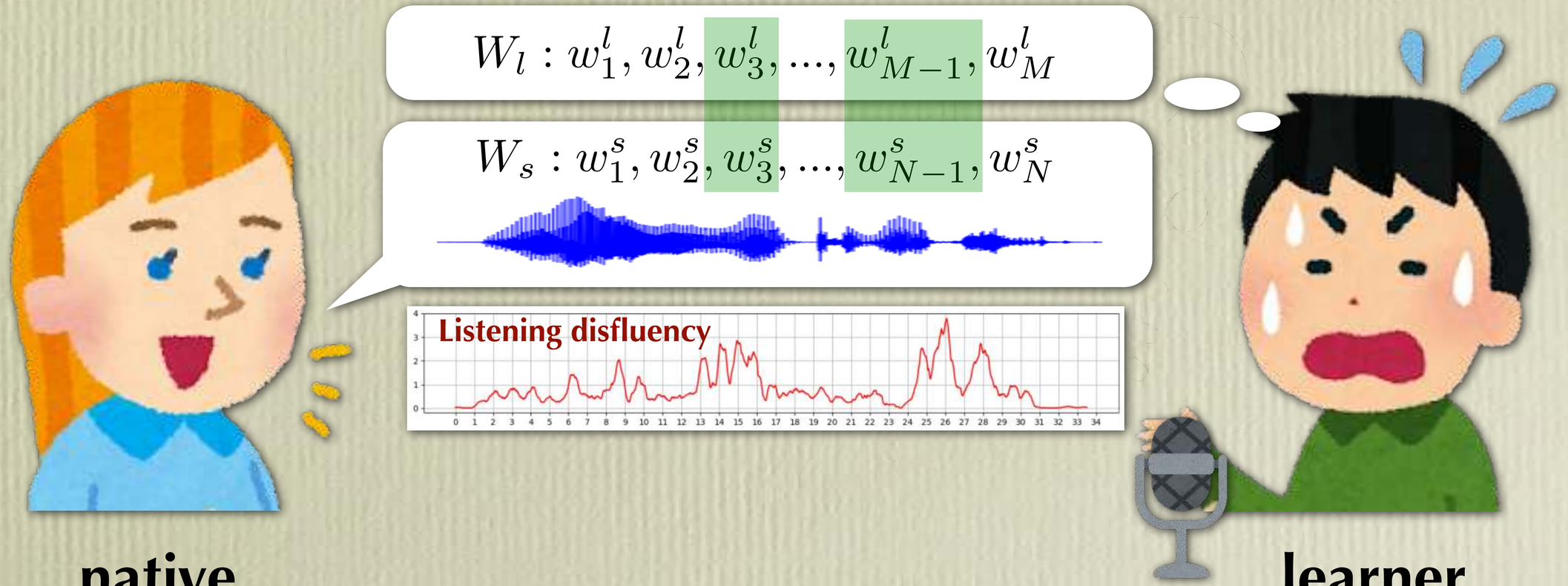


Inoue, Y., et al. (2018) A Study of Objective Measurement of Comprehensibility through Native Speakers' Shadowing of Learners' Utterances. Proc. Interspeech 2018, 1651-1655

Assessment of listening [Inoue+'18]

When listening, where in a given speech does LD take place?

- Listening is mental activity and does not present any acoustic events.

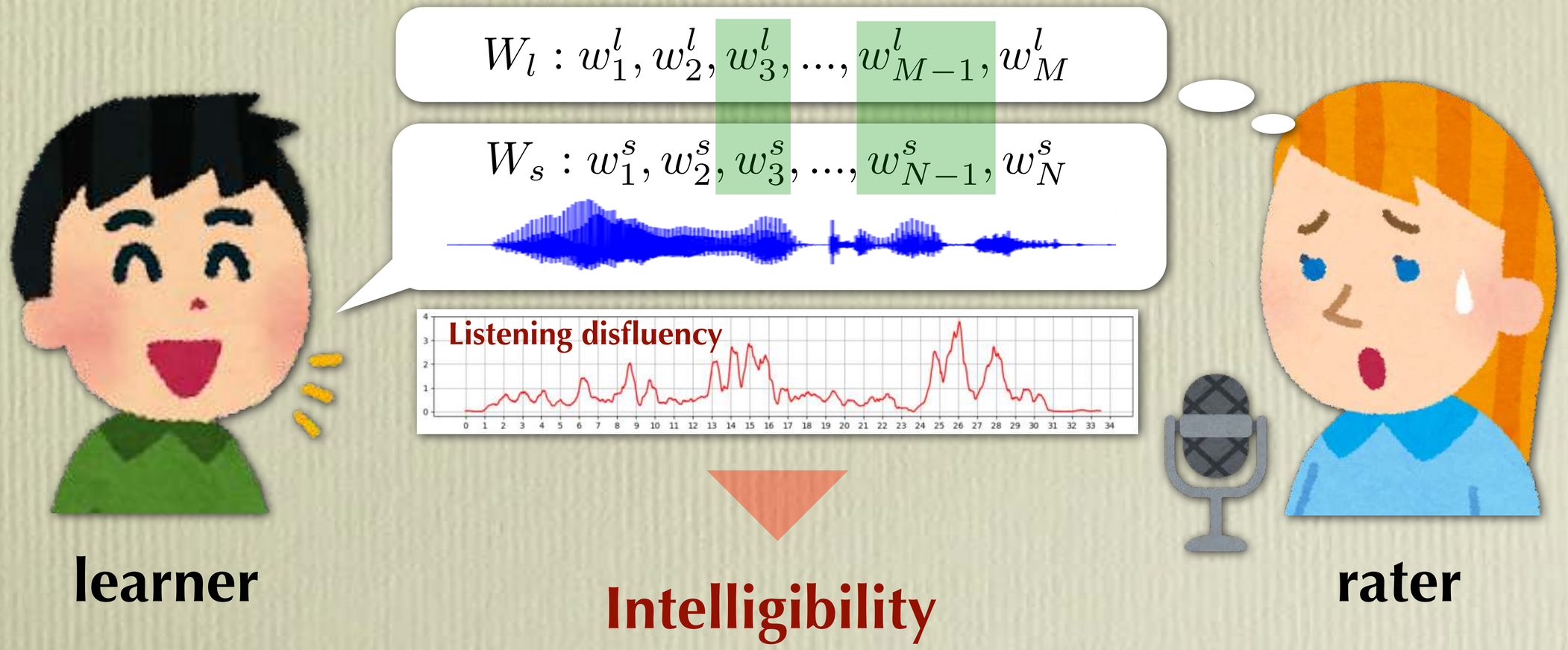


Inoue, Y., et al. (2018) A Study of Objective Measurement of Comprehensibility through Native Speakers' Shadowing of Learners' Utterances. Proc. Interspeech 2018, 1651-1655

Assessment of speaking based on listening [Inoue+'18]

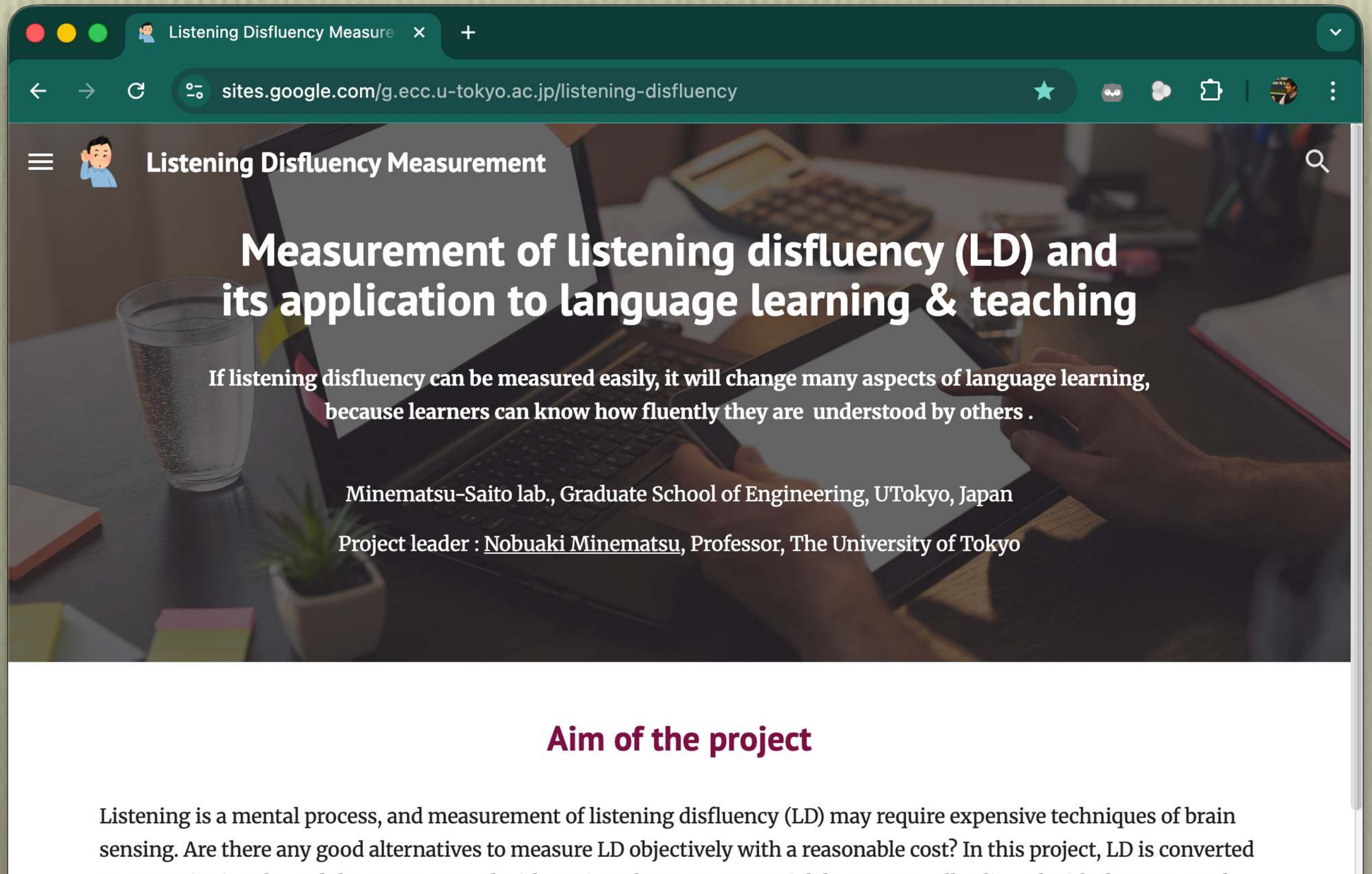
When listening, where in a given speech does LD take place?

- Listening is mental activity and does not present any acoustic events.



Inoue, Y., et al. (2018) A Study of Objective Measurement of Comprehensibility through Native Speakers' Shadowing of Learners' Utterances. Proc. Interspeech 2018, 1651-1655

The web site of our project



The web site of our project

Listening Disfluency Measure

sites.google.com/g.ecc.u-tokyo.ac.jp/listening-disfluency/home?authuser=0

Listening Disfluency Measurement

How to measure LD?

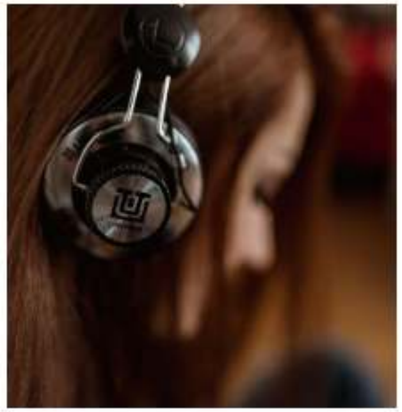
Shadowing-based acoustic measurement of LD

Shadowing and mirroring

Measurement of learners' LD

To assess learners' listening skills

To reduce learners' listening disfluency



Measurement of raters' LD

To assess learners' speaking skills

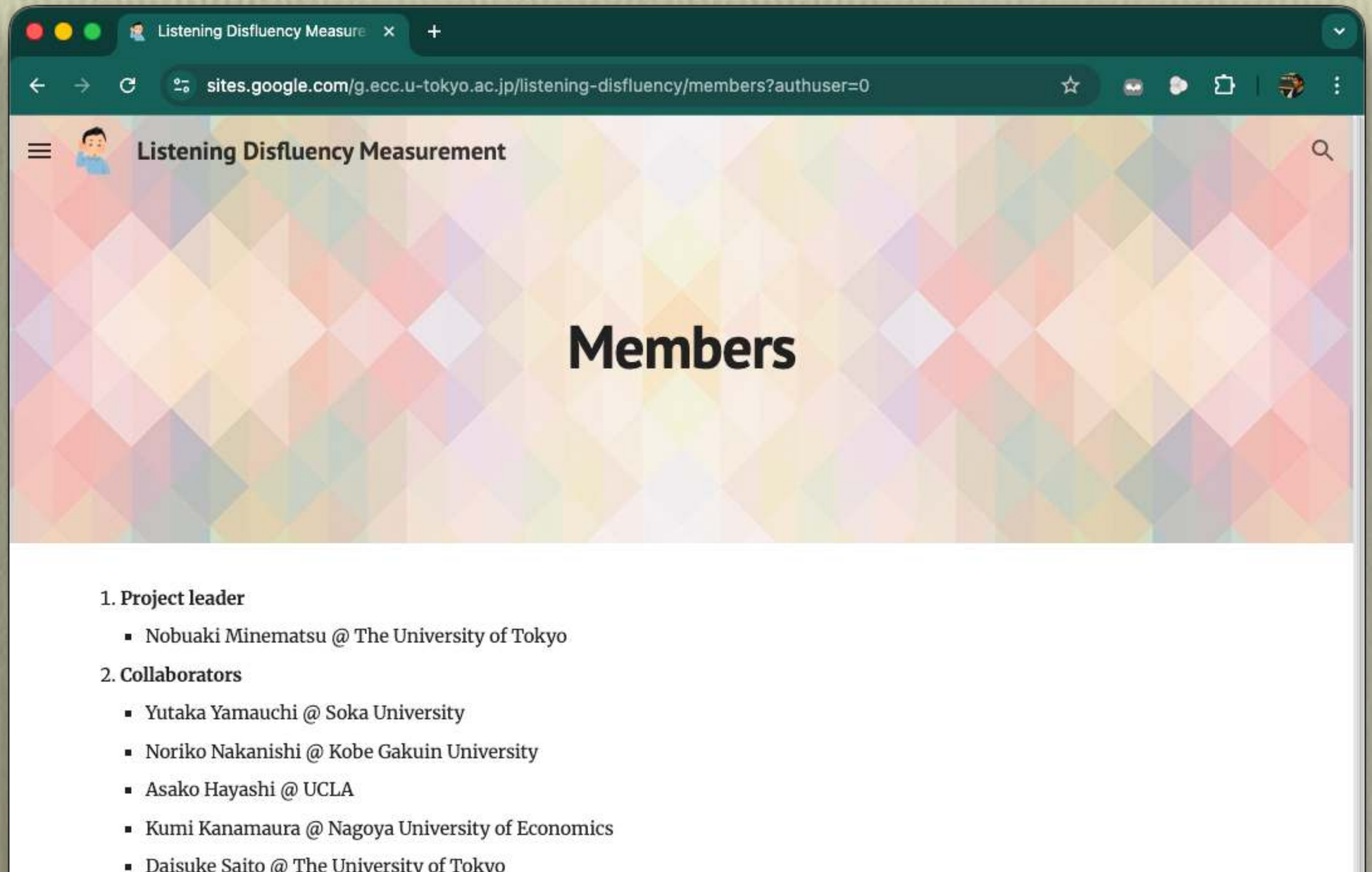
To score the intelligibility for learners' speech

Prediction of LD

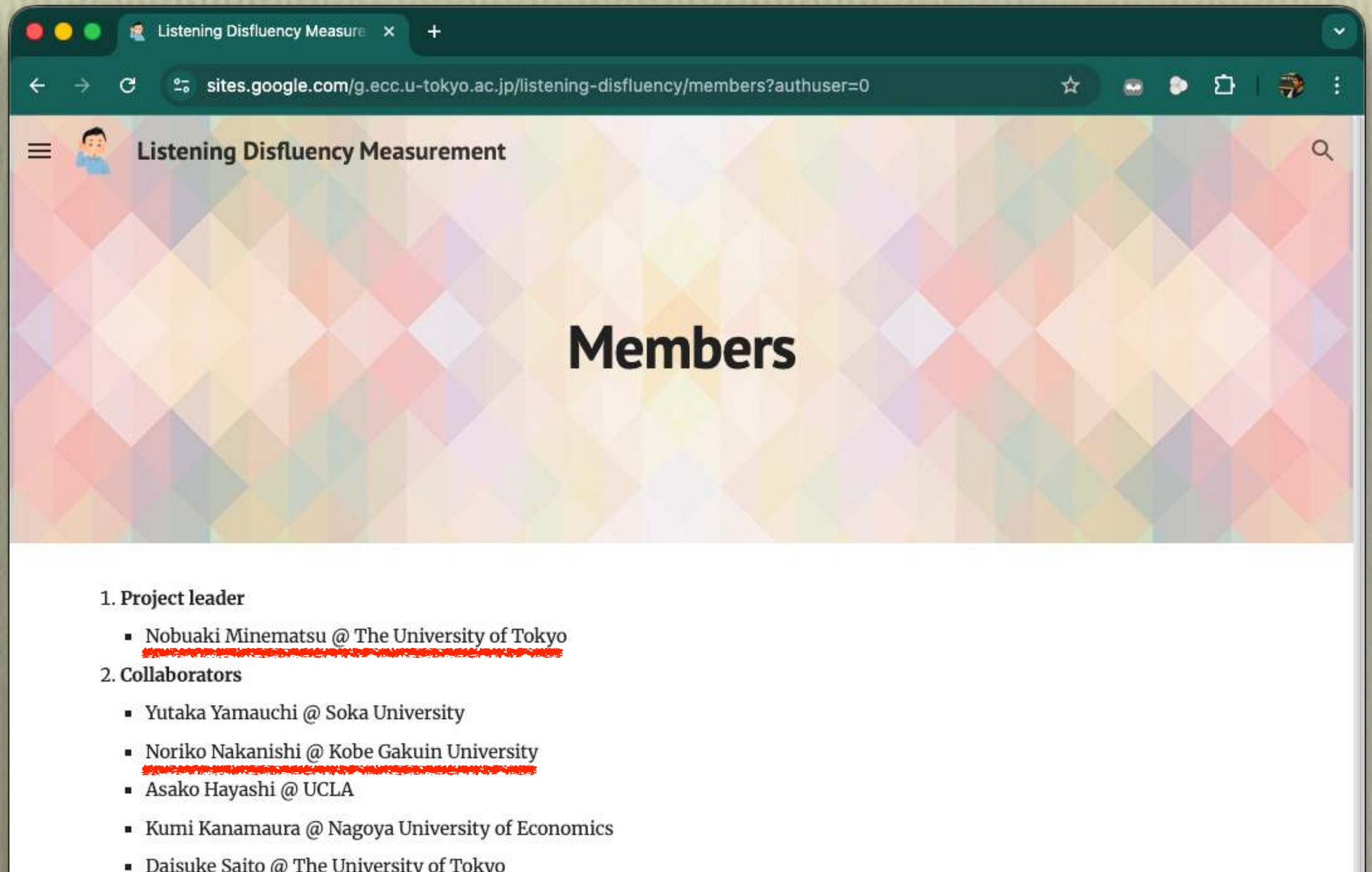
Prediction of learners' LD using machine learning

Sequence-to-sequence voice conversion is applied to predict raters' LD.

The web site of our project

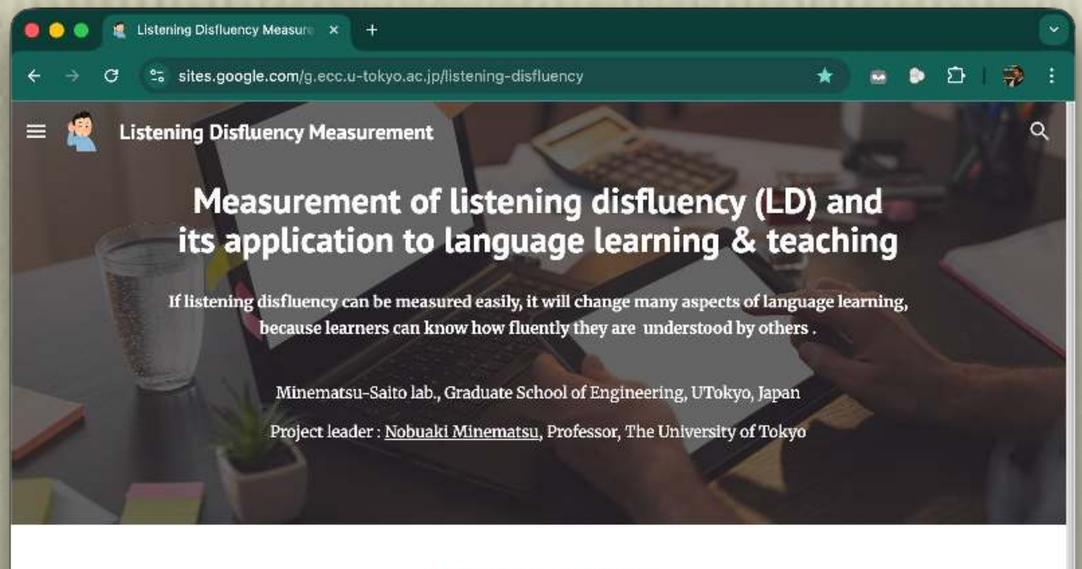


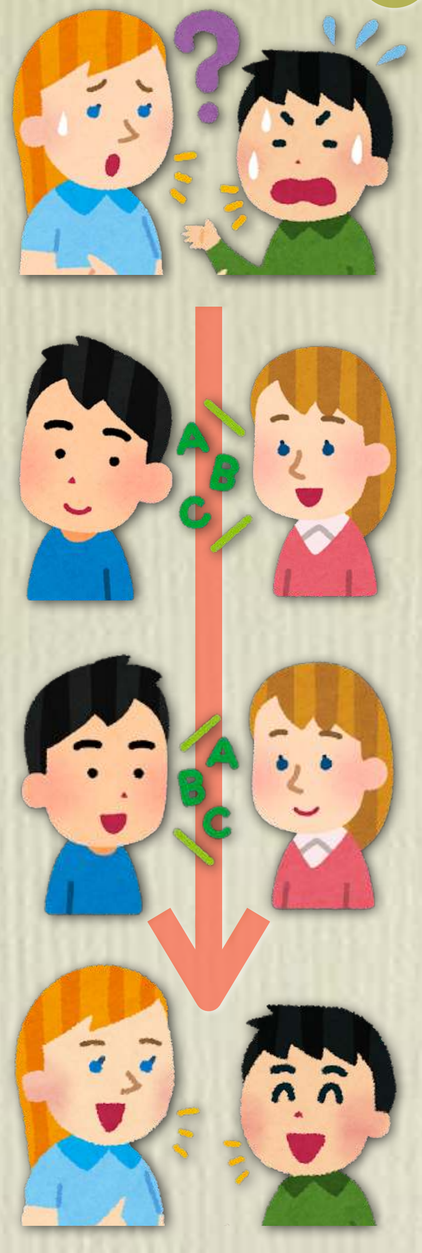
The web site of our project



Outline of this talk

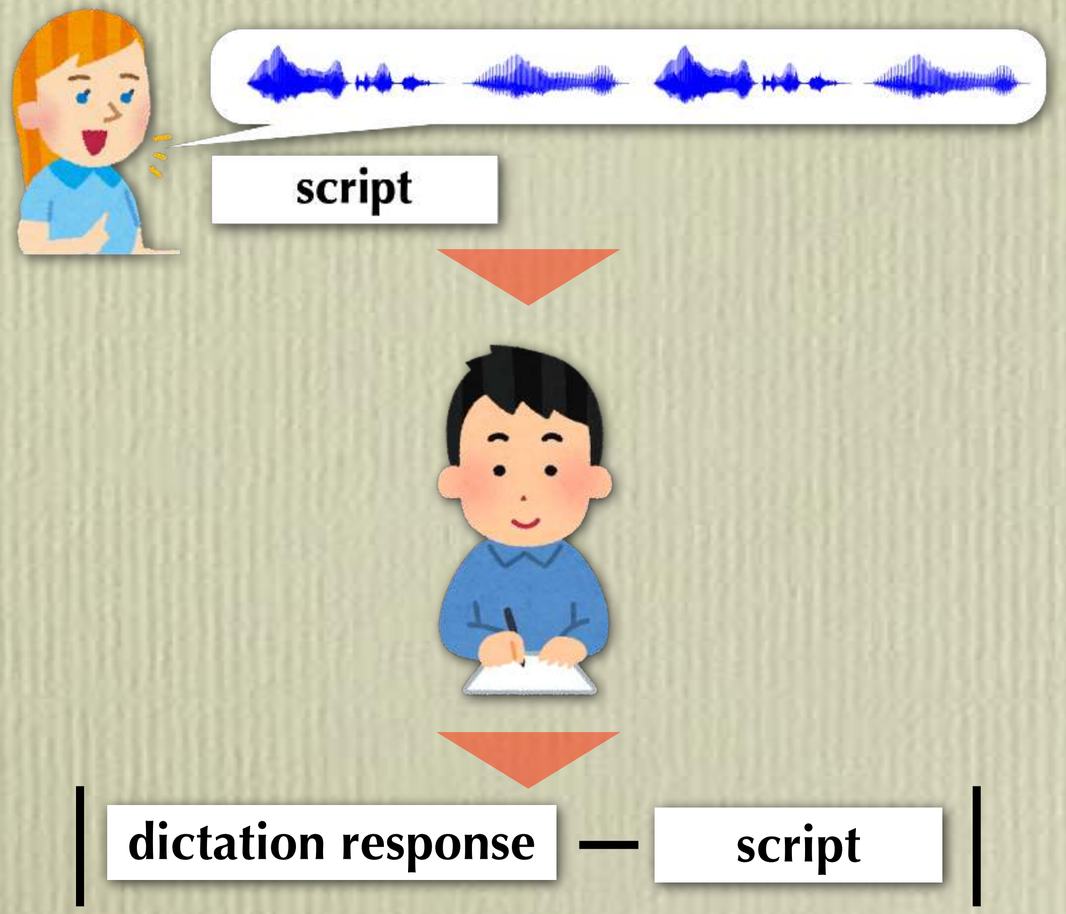
- Why listening disfluency (LD)?
- How to measure LD with a microphone?
- Measurement and analysis of learners' LD
- Measurement and analysis of raters' LD
- Prediction of raters' LD
- Conclusions





Measurement of listening in class

Dictation test by writing using a hand

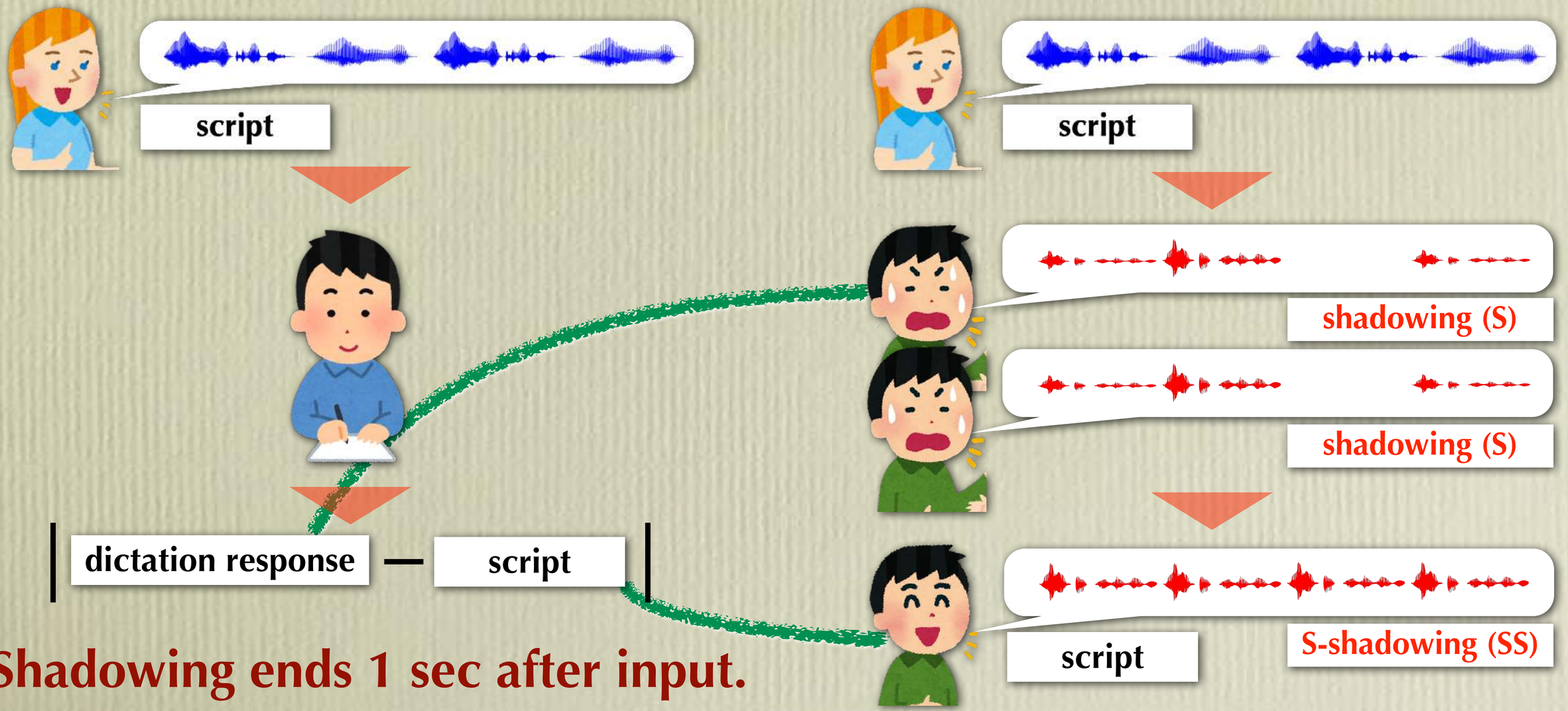


- A long utterance cannot be dictated only with one-time listening [Munro+'20].
  - Memory capacity is limited and short utterances are only available.
- Listeners may be dictating with deep guessing [Thomson'18].
  - Listeners might reconstruct and rephrase what they actually heard.
- Listeners may struggle to recall the orthography of perceived words [Minematsu+'21].
  - Is this test really a listening test?

Writing using a hand takes long.

Measurement of listening in class

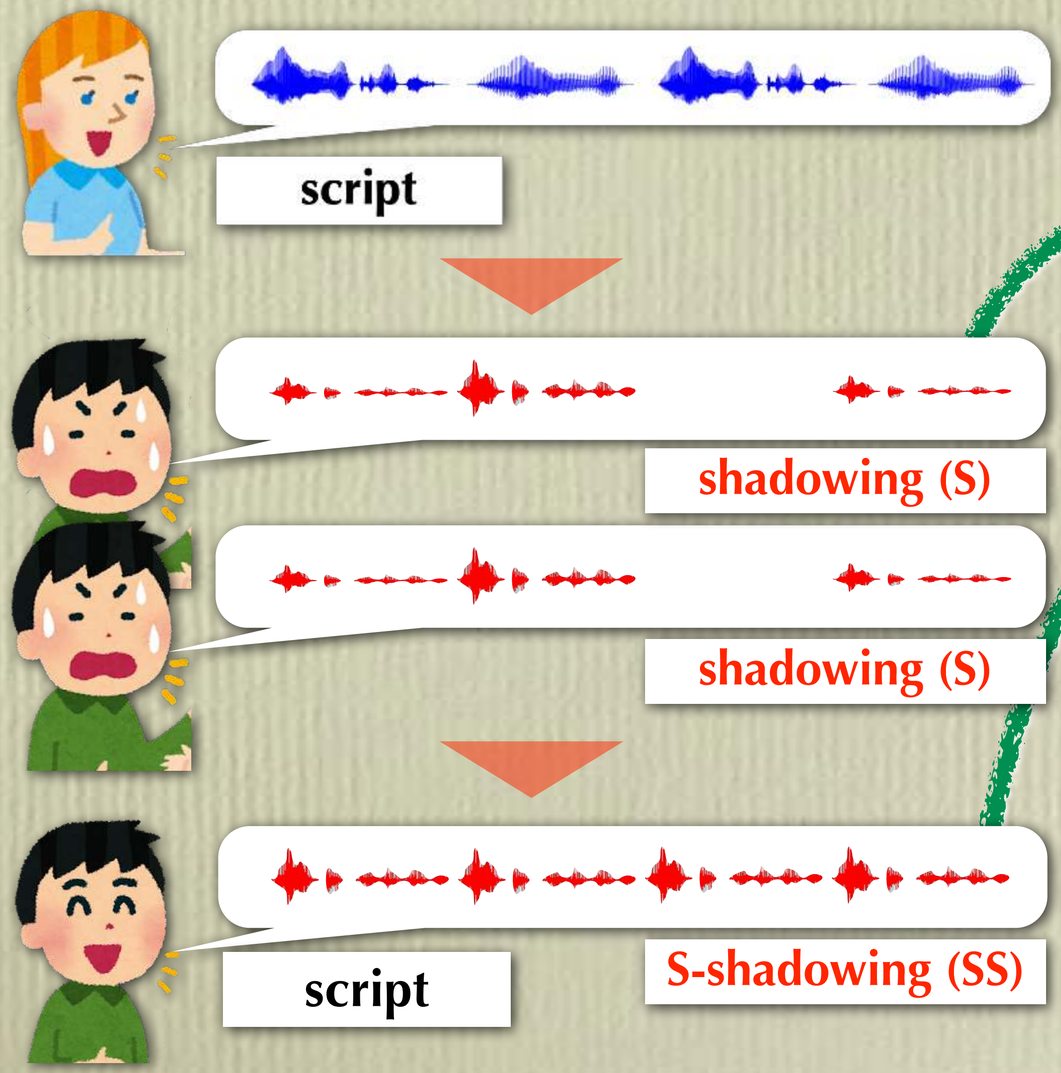
Dictation test by speaking using a mouth [Inoue+'18, Zhu+'20]

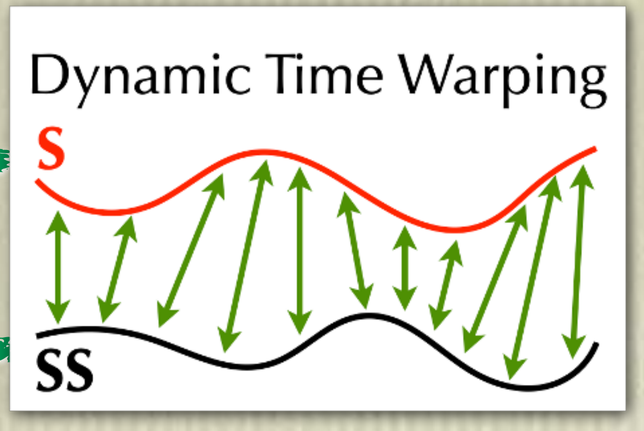


Shadowing ends 1 sec after input.

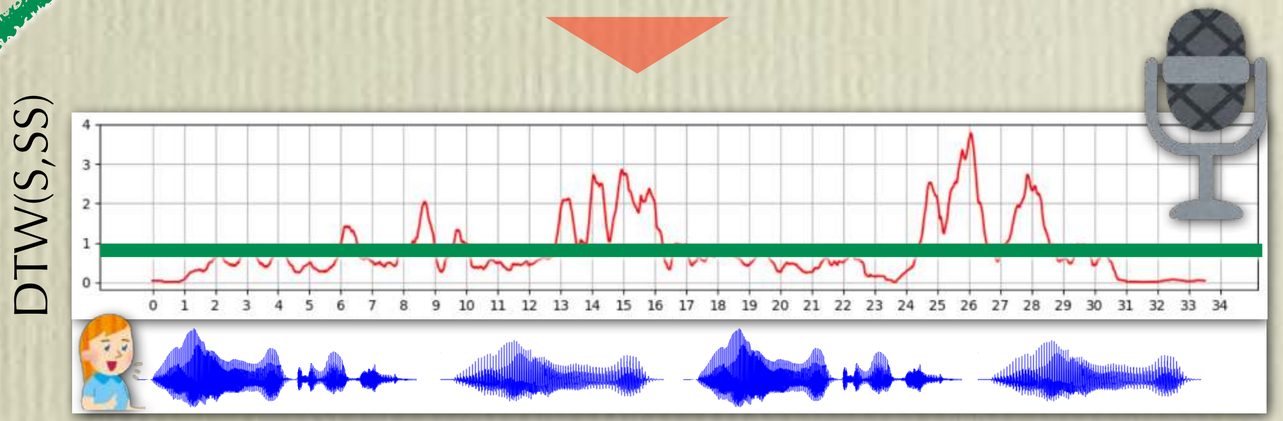
Measurement of listening in class

Dictation test by speaking using a mouth [Inoue+'18, Zhu+'20]





PPG-DTW
(**P**honetic
Posterior**G**ram)

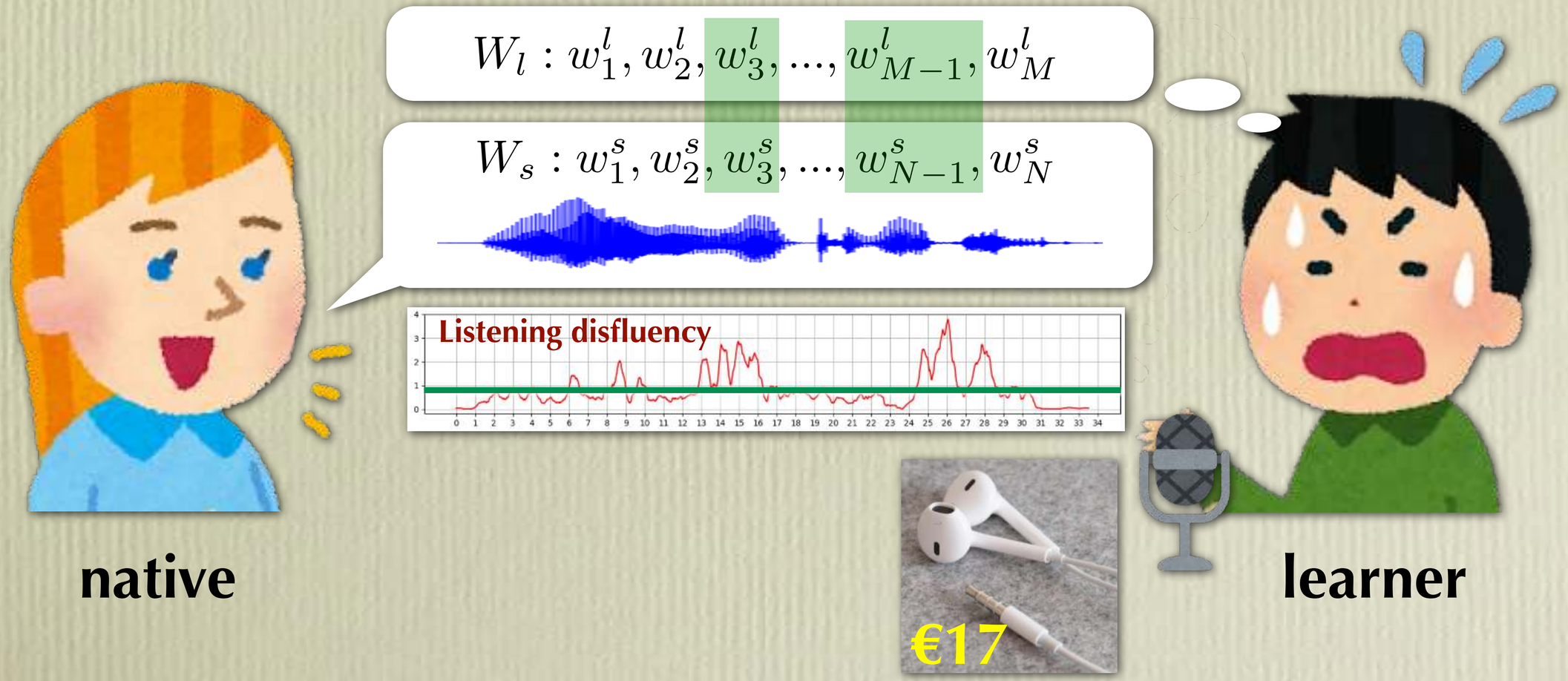


Shadowing disfluency = listening disfluency

Assessment of listening

When listening, where in a given speech does LD take place?

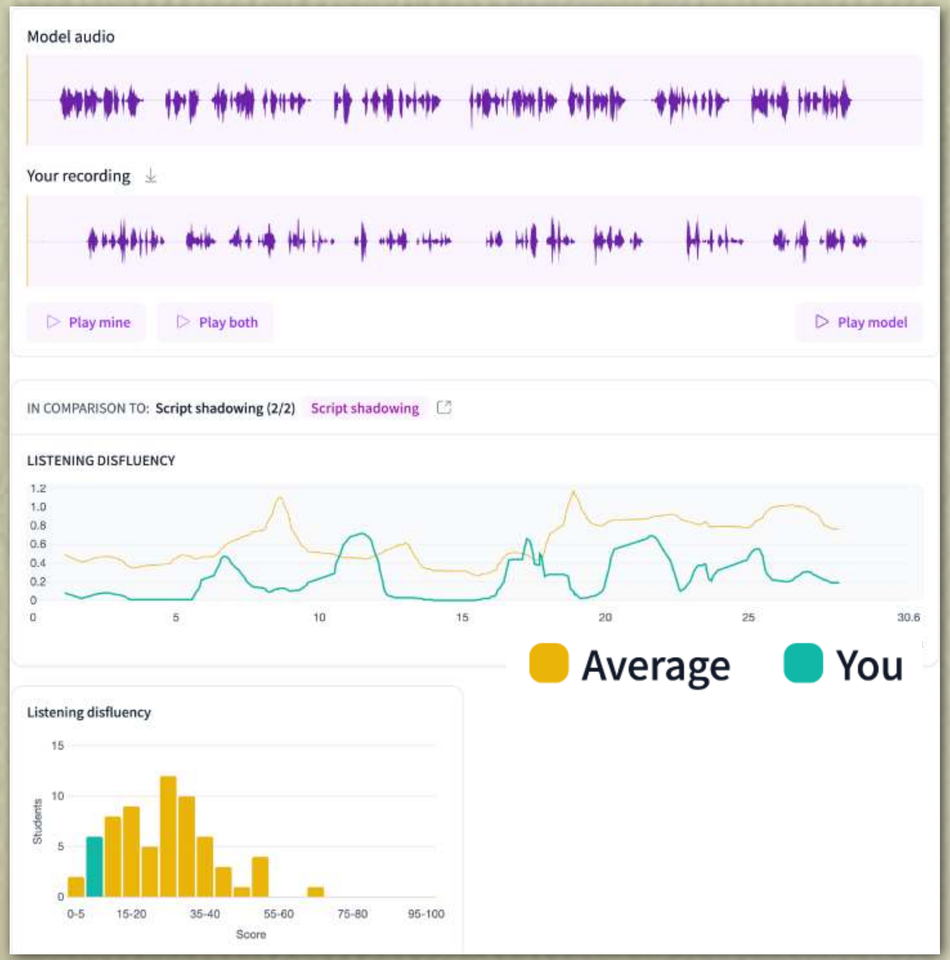
- Listening is mental activity and does not present any acoustic events.

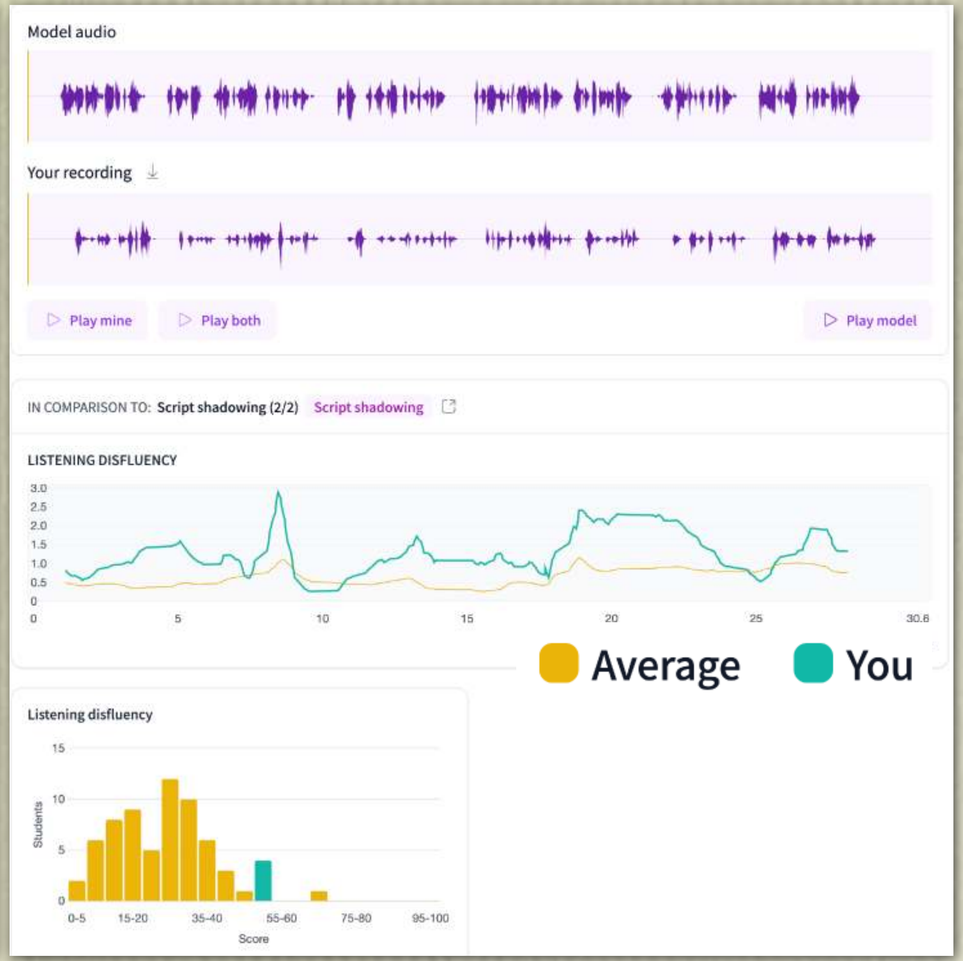


Special Training for English Academic Communication

Two-month intensive daily course for comprehensive speech training

- For listening, pronunciation, and conversation with ChatGPT
- About 30-min daily practices during a summer or spring break





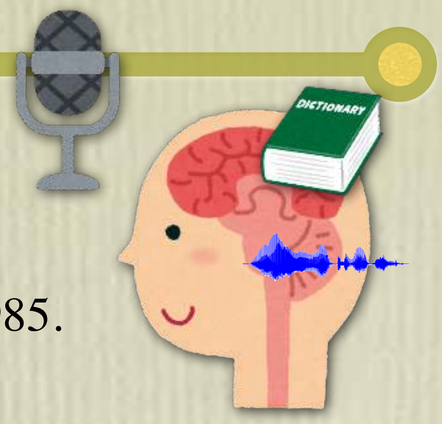
聞ける耳。伝わる口。考える頭。

STEACは，日頃英語の音に接していない耳と口と頭を英語漬けにすることを狙った，夏／春休み毎日30分のオンデマンド特訓授業です。音声技術・言語技術，そしてAI技術を用いて，皆さんの「聞く」「話す」「考える」を鍛え，皆さんの能力を可視化し，スコア化し，評価し，その都度，フィードバックを返します。

夏休み：工学部３年生対象，春休み：工学部２年生対象（各１単位）

工学系及び情報理工学系研究科は2026年度から授業が英語化されます
学部生のうちに「聞ける耳，伝わる口，考える頭」を身につけましょう

History of shadowing



Originally proposed for psycho-linguistic studies in 1980s

- To measure and analyze the human process of listening acoustically

  Marslen-Wilson, W. D. “Speech shadowing and speech comprehension,” Speech Comm., 4, 55-73, 1985.

Subsequently applied to language training in 1990s

- To train simultaneous interpreters to enhance the capacity of their working memory

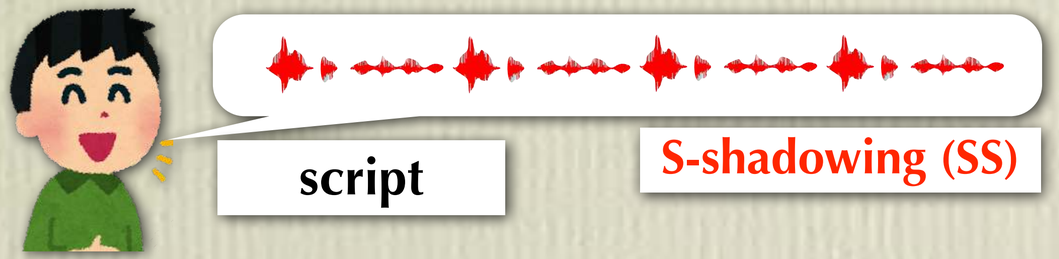
  Kurz, I. ”Shadowing exercises in interpreter training,” Teaching Translation and Interpreting: Training, Talent and Experience, Cay Dollerup and Anne Loddegaard (eds.), John Benjamins, 1992.

- To train language learners to improve their listening skills

  Tamai, K. ”The effectiveness of shadowing on listening skill improvement and its’ interpretation in the context of language education," Current English Studies 36, pp.105-16, 1997

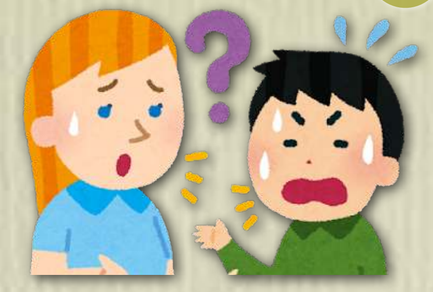
Reintroduced again to language training as analysis tool in 2020

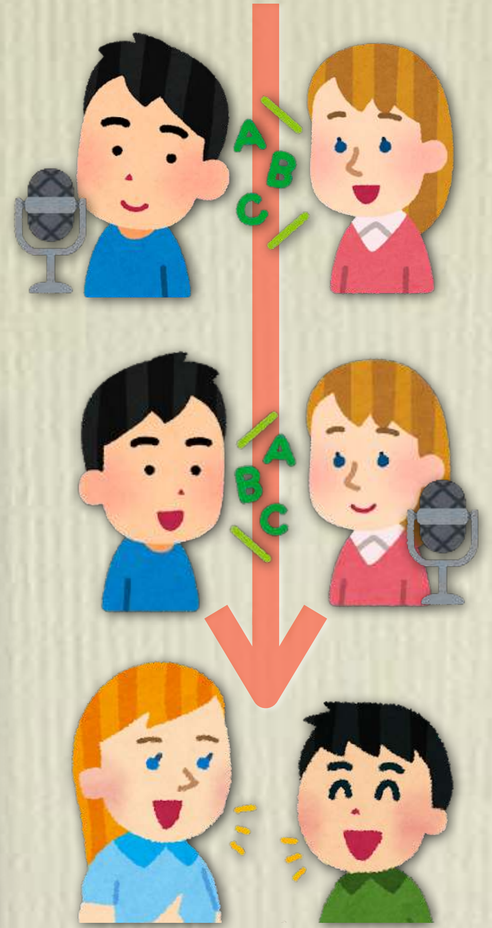
- To quantify listening disfluency in sequence by comparing shadowing and s-shadowing

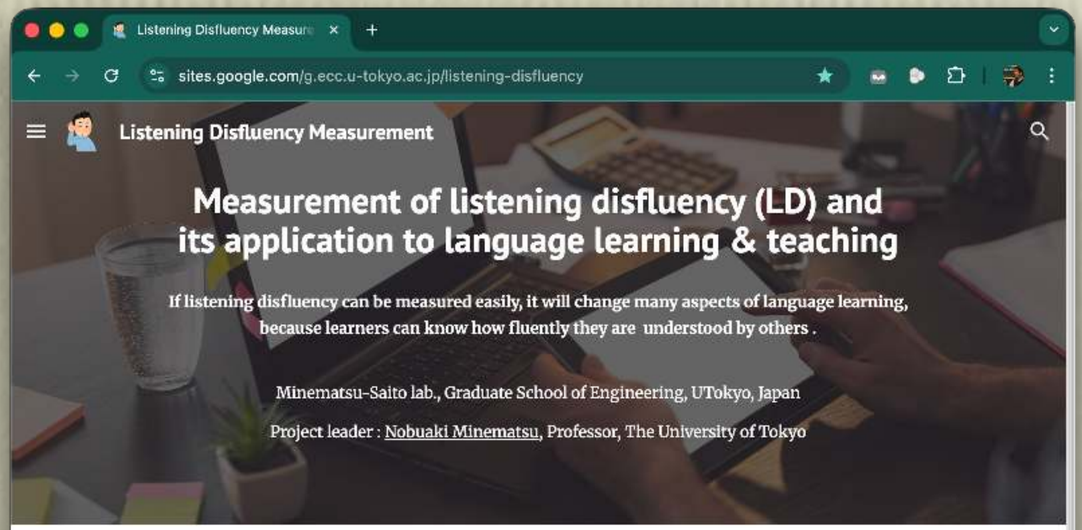


Outline of this talk

- Why listening disfluency (LD)?
- How to measure LD with a microphone?
- Measurement and analysis of learners' LD
- Measurement and analysis of raters' LD
- Prediction of raters' LD
- Conclusions







Shadowing marathon in 2021 [Kunihara+'22]

42-day (6-week) intensive training of shadowing

- Participants: 35 beginning and intermediate learners of English
  - L1 = Japanese, CEFR level = A1 and A2
- Task: **S1 - S2 - S3 - SS - R** conducted for a model utterance (**M**)
  - No corrective feedback on learners' pronunciation is provided.

Model
audio

PPG-DTW(S1,SS)

Shadow1
audio only

Shadow2
audio only

Shadow3
audio only

Script-shadow
audio + text

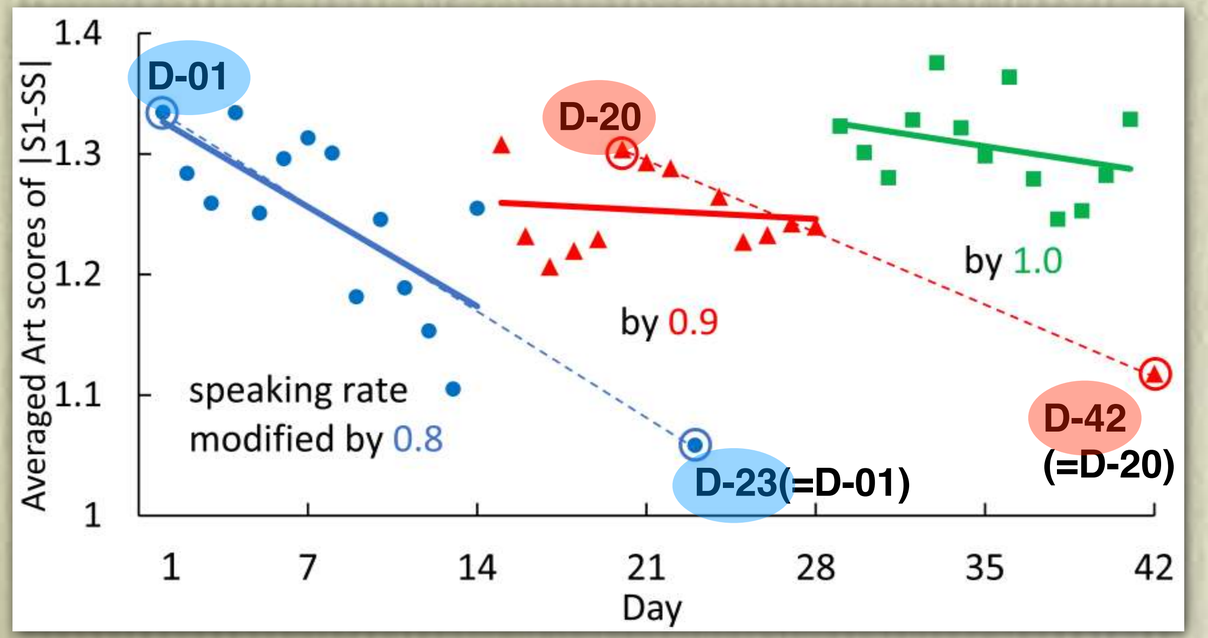
Reading-aloud
text only

Kunihara, T., et. al., (2022) Gradual Improvements Ob… …tion of L2 Sounds Through Continuing Shadowing Practices on a Daily …

Shadowing marathon in 2021 [Kunihara+'22]

42-day (6-week) intensive training of shadowing

- Participants: 35 beginning and intermediate learners of English
  - L1 = Japanese, CEFR level = A1 and A2
- Task: **S1 - S2 - S3 - SS - R** conducted for a model utterance (**M**)
  - No corrective feedback on learners' pronunciation is provided.
  - M = a monologue utterance of about 30-sec for A1/2 level.
  - 4 model utterances / day x 42 days = 168 model utterances in total.



Model
audio

Shadow1
audio only

Shadow2
audio only

Shadow3
audio only

Script-shadow
audio + text

Reading-aloud
text only

PPG-DTW(S1,SS)

Diversity of English pronunciations

Whose English is easy or difficult for you to understand?
Who finds your English to be easy or difficult to understand?

- Diversity of listening behaviors as well as that of speaking behaviors
- Listening difficulty is directional. How to measure the directional difficulty?

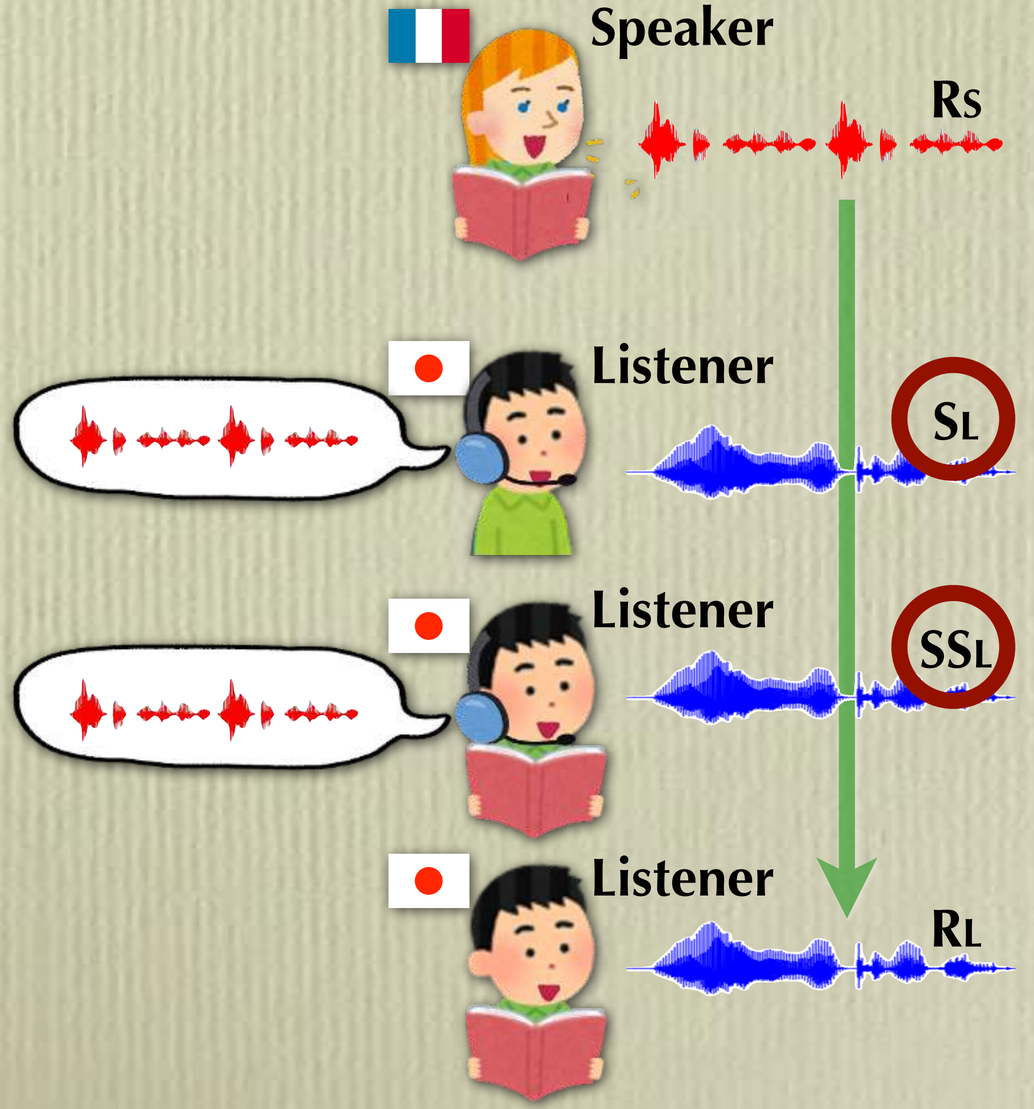
How to visualize the directional characteristics of listening?

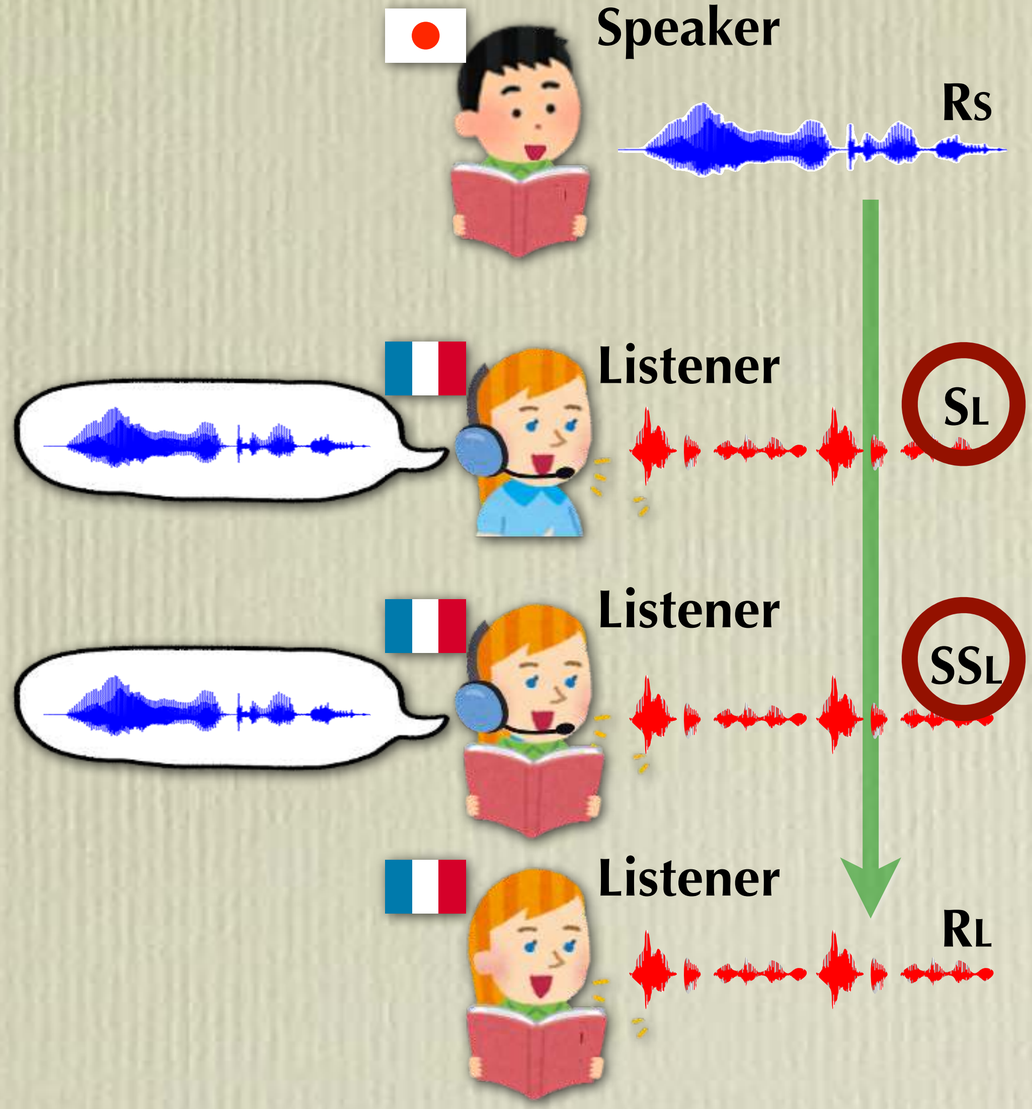
- X listens to various others, who listen to X.
- Visualization of X's communicability with various others using **C**-chart

Tomita, Y., et al. (2024) Analysis and Visualization of Directional Diversity in Listening Fluency of World Englishes Speakers in the Framework of Mutual Shadowing. Proc. Interspeech 2024, 4024-4028

Listening Disfluency (LD) and Pronunciation Gap (PG)

A French talks with a Japanese.



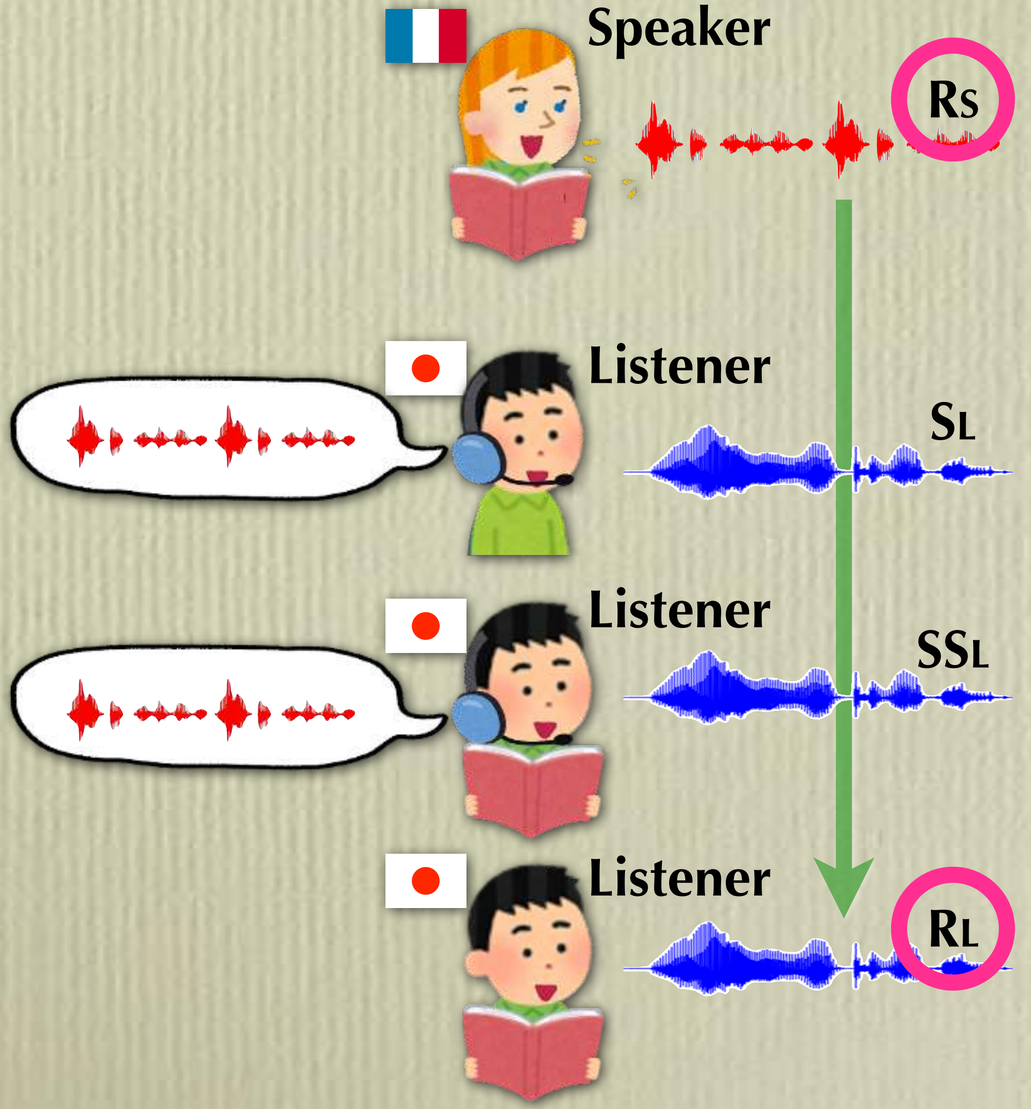


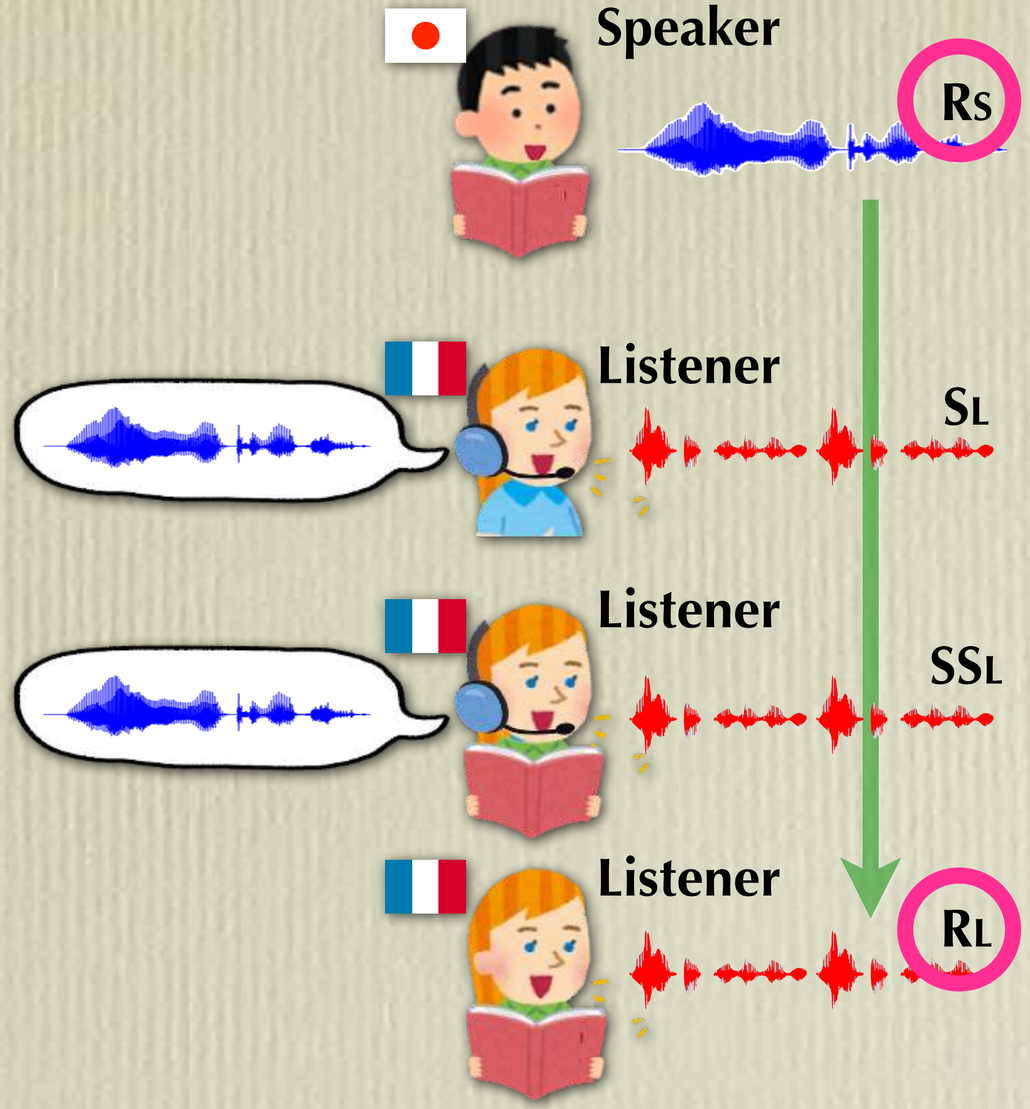
LD = DTW(SL, SSL)

PG = DTW(RS, RL)

Listening Disfluency (LD) and Pronunciation Gap (PG)

A French talks with a Japanese.





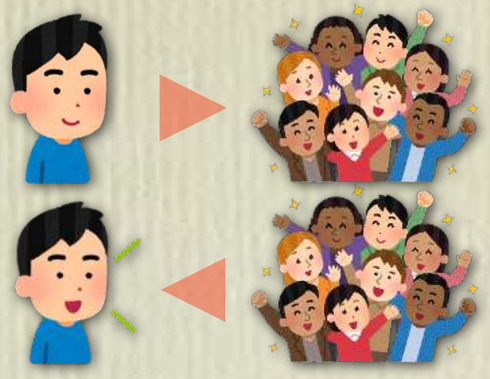
LD = DTW(Sl, SSl)

PG = DTW(Rs, Rl)

Mutual shadowing among WE speakers [Tomita+'24]

28 Participants in the mutual shadowing experiment

- Recruited from Japanese classes at UTokyo and divided into 3 groups with good care paid to the diversity of language background of each group's participants.
- 28 passages with 30-sec length were selected from Eiken Grade-2 Test (ARI = 6.2 to 7.0).
- A mutual shadowing experiment was conducted for each group.
  - Everybody shadowed everybody, and everybody was shadowed by everybody.



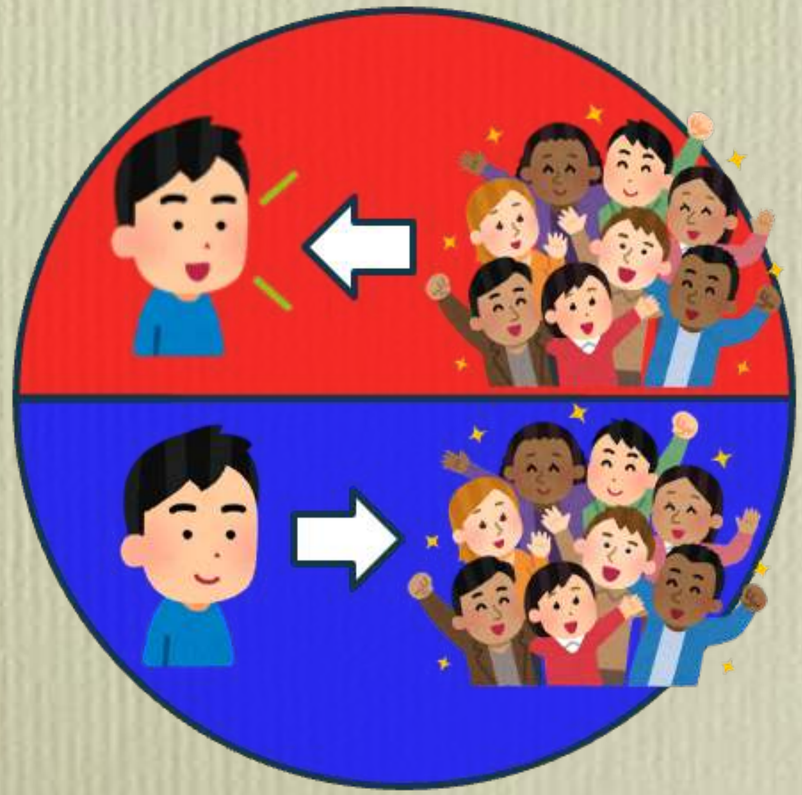
	group A				group B				group C		
ID	lang.	family	C-chart	ID	lang.	family	C-chart	ID	lang.	family	C-chart
1	CHN	ST	e	1	CHN	ST	b̄,e,f	1	CHN	ST	c̄,e
2	CHN	ST	a,e	2	CHN	ST	d̄	2	CHN	ST	d̄,e
3	JPN	TU	a,e	3	JPN	TU	ā,f	3	JPN	TU	b̄,e,f
4	KOR	TU	a,e	4	KOR	TU	b̄,e	4	KOR	TU	b̄,e
5	KOR	TU	a,e	5	FRA	IE (IT)	c̄,e	5	FRA	IE (IT)	b̄,f
6	FRA	IE (IT)	d,e	6	ITA	IE (IT)	c̄,e	6	SPN	IE (IT)	d,f
7	ITA	IE (IT)	c̄,f	7*	HIN	IE (II)	d	7*	HIN	IE (II)	b,e
8*	HIN	IE (II)	d,f	8	UKR	IE (SL)	c̄,f	8	UKR	IE (SL)	f
9	SRB	IE (SL)	c,e	9*	MAL	DR	d	9	HUN	UR (FU)	c
10	HUN	UR (FU)	c,e								

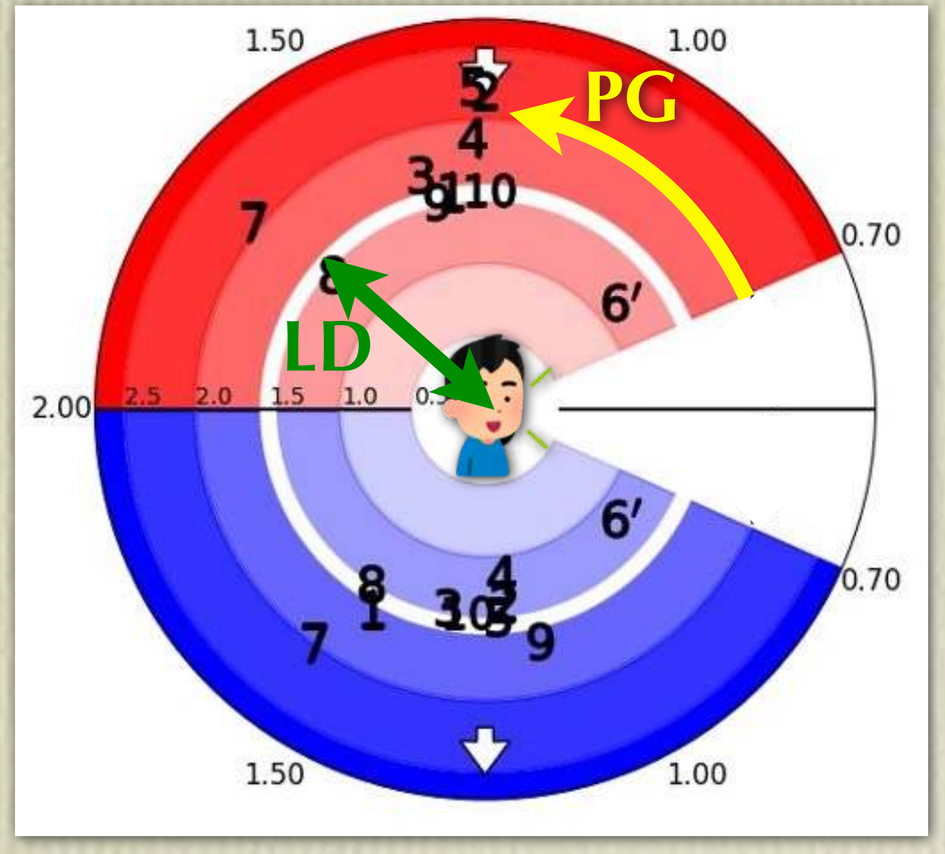
languages		lang. families	
CHN	Chinese	ST	Sino-Tibetan
JPN	Japanese	TU	Trans-Eurasian
KOR	Korean	IE	Indo-European
FRA	French	UR	Uralic
ITA	Italian	DR	Dravidian
SPN	Spanish		
HIN	Hindi	lang. sub-families	
SRB	Serbian	IT	Italic
UKR	Ukrainian	II	Indo-Iranian
HUN	Hungarian	SL	Slavic
MAL	Malayāḷam	FU	Finno-Ugric

Simultaneous visualization of various LDs and PGs

C-chart to visualize global communicability of a specific speaker

- Visualization of **how fluently the WE speakers listen to a specific speaker ( )** and **how fluently that speaker listens to the WE speakers.**

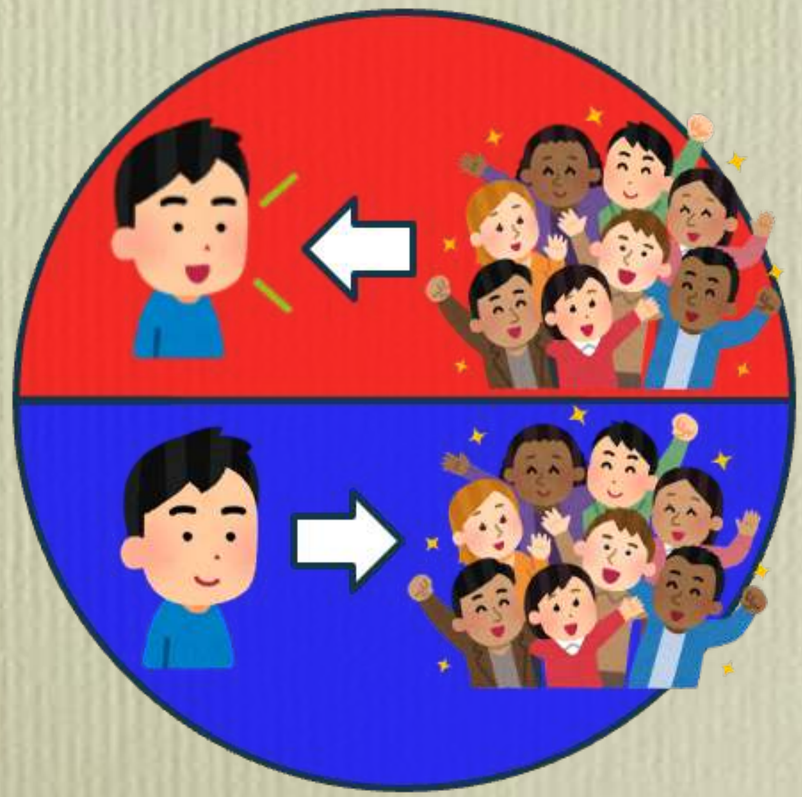


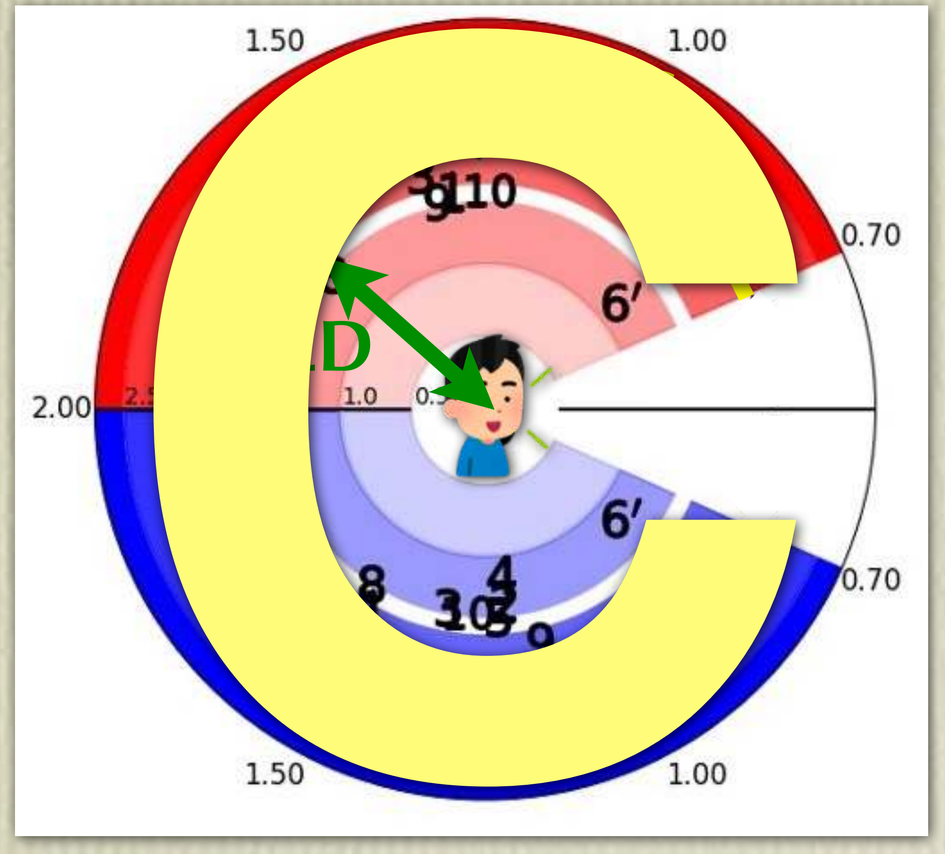


Simultaneous visualization of various LDs and PGs

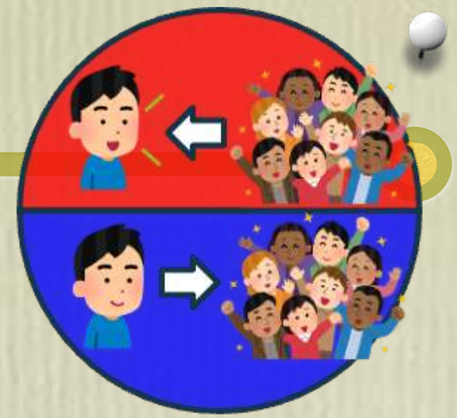
C-chart to visualize global communicability of a specific speaker

- Visualization of **how fluently the WE speakers listen to a specific speaker ( )** and **how fluently that speaker listens to the WE speakers.**

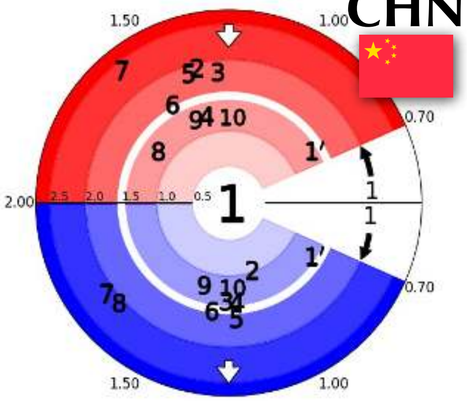




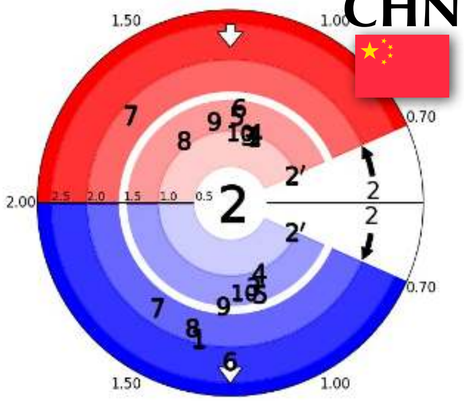
C-charts of the participants in group A



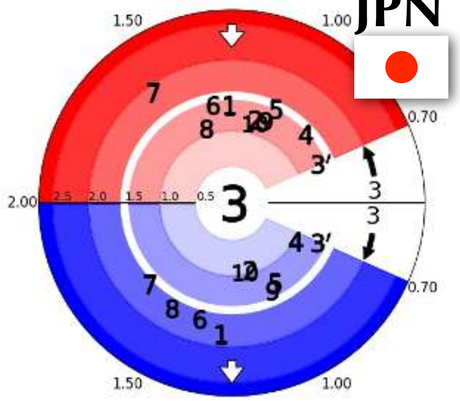
CHN



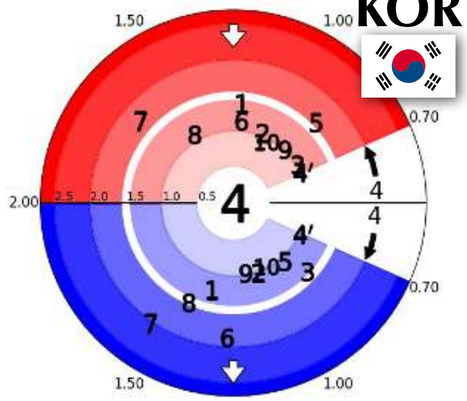
CHN



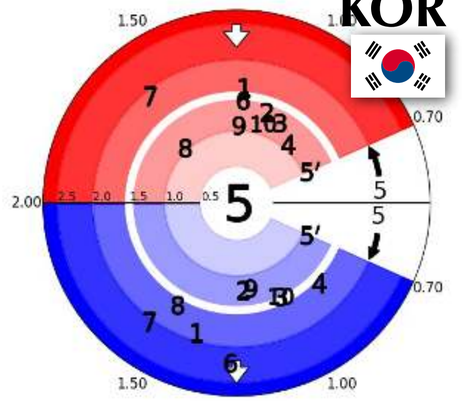
JPN



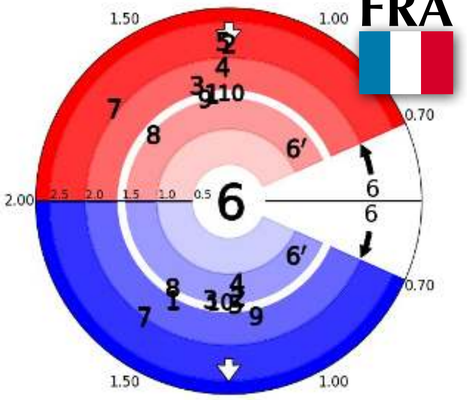
KOR



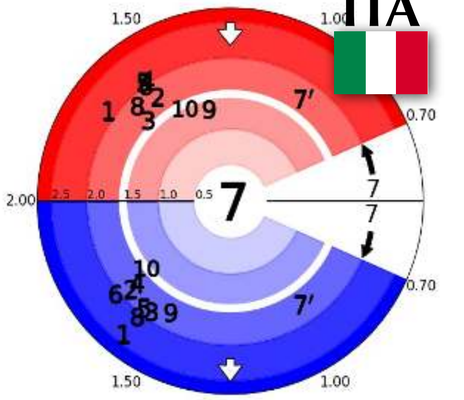
KOR



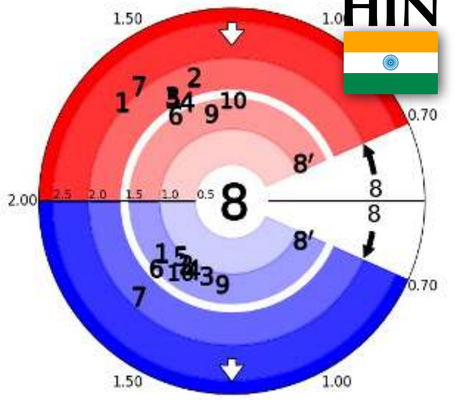
FRA



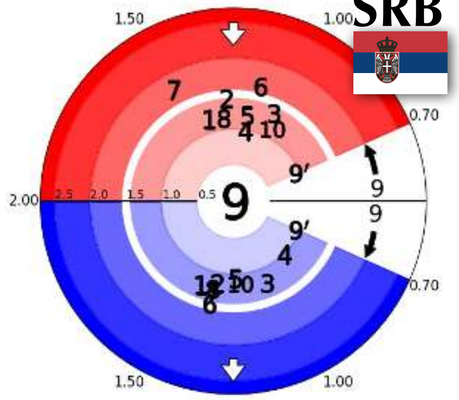
ITA



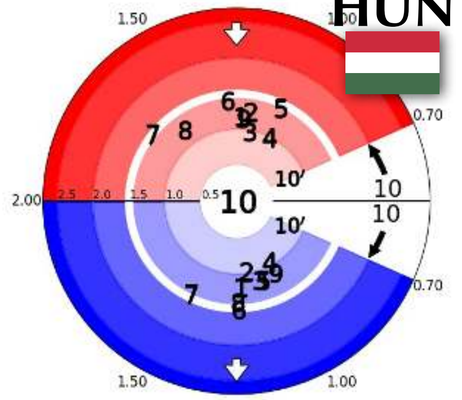
HIN



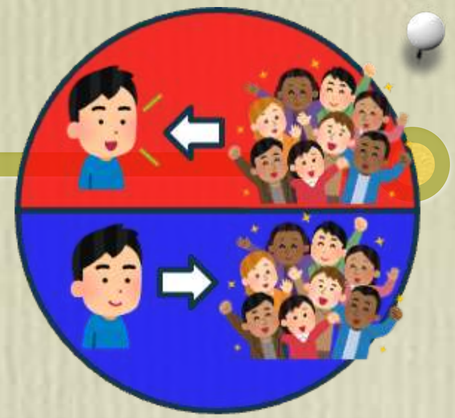
SRB

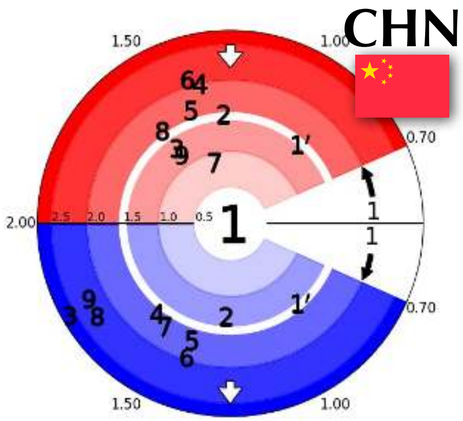


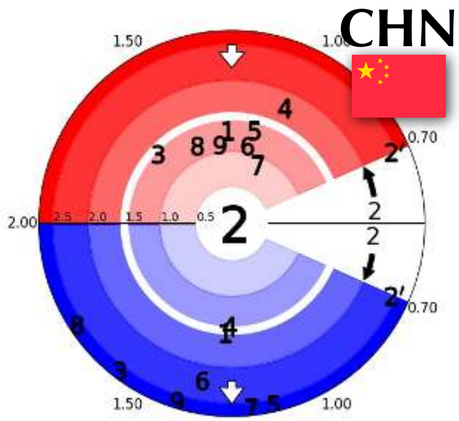
HUN

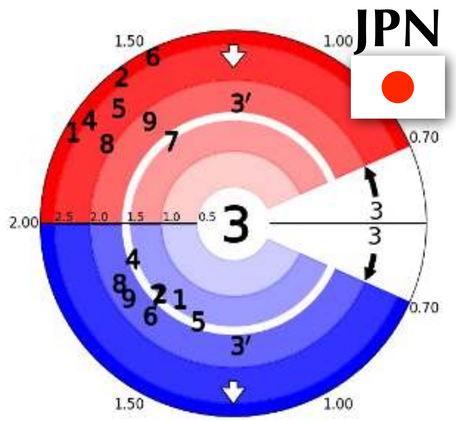


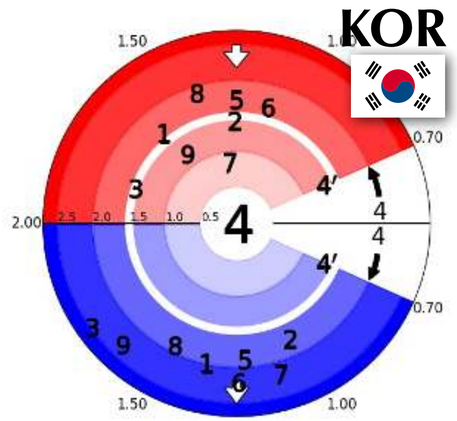
C-charts of the participants in group B

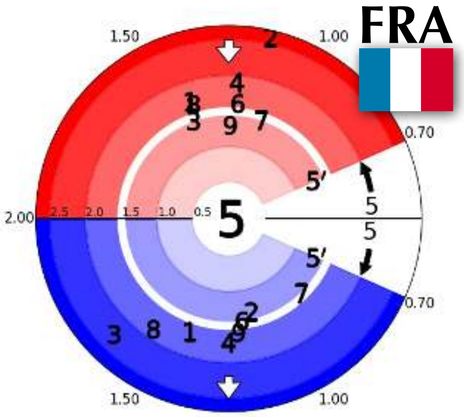


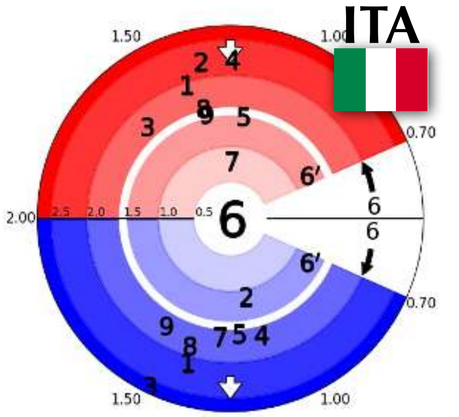


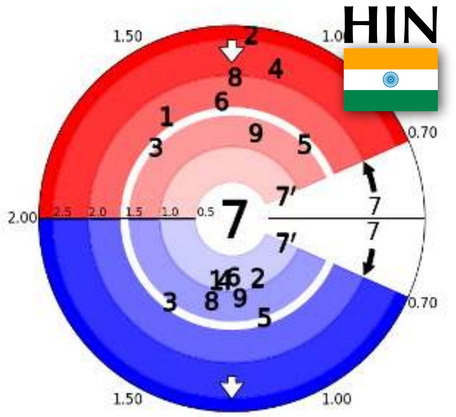


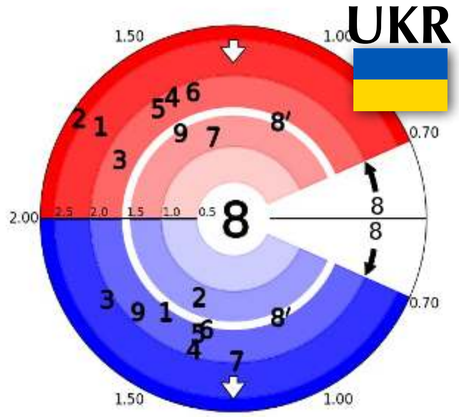


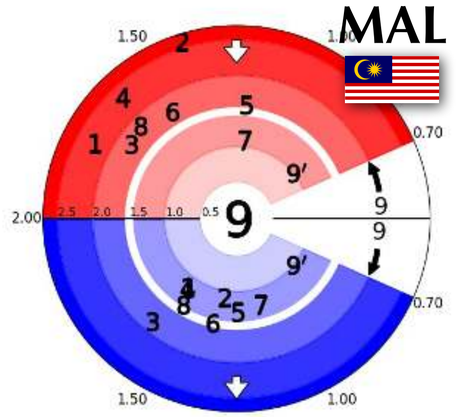




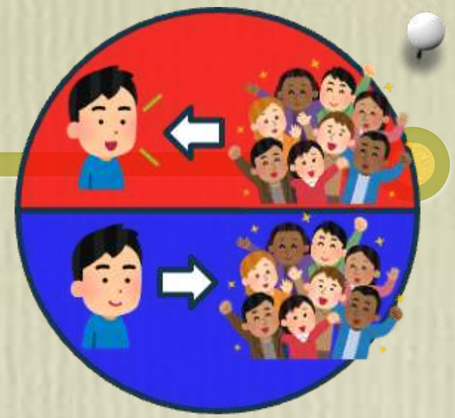


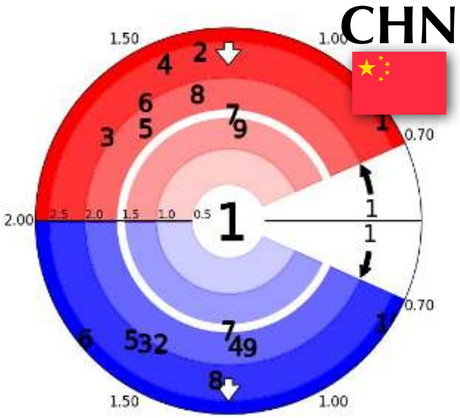


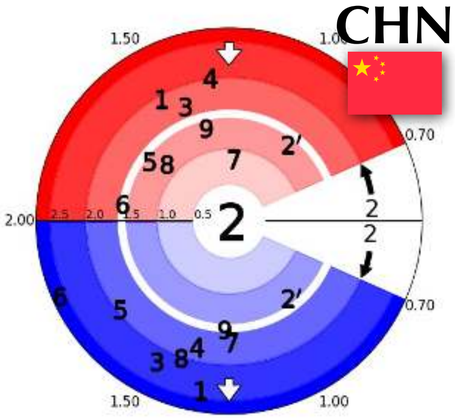


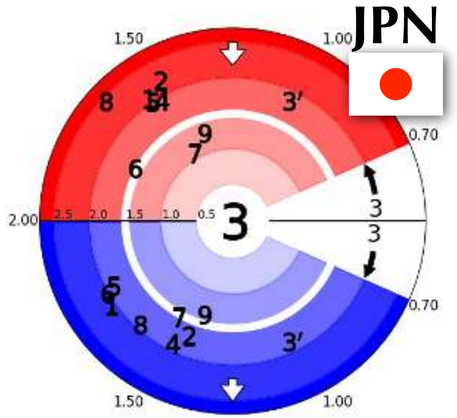


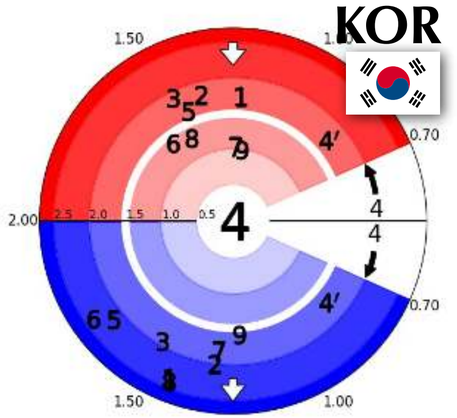
C-charts of the participants in group C

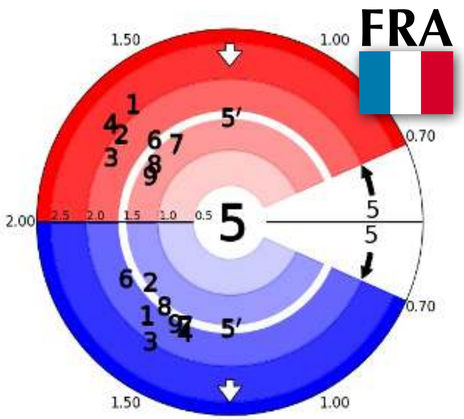


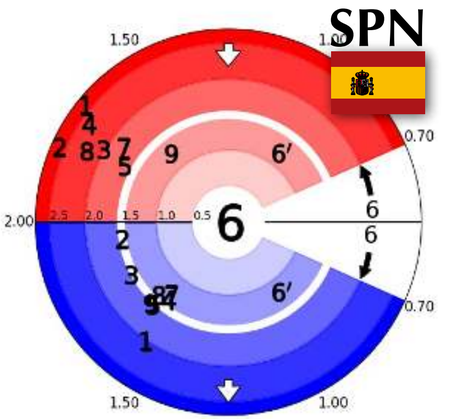


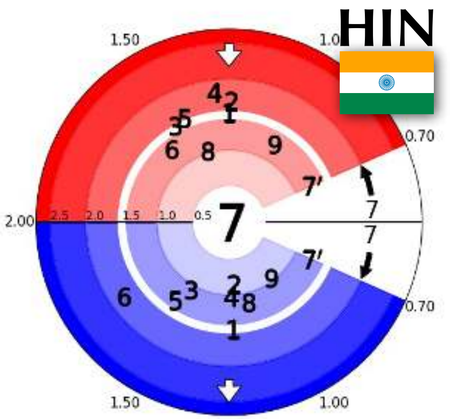


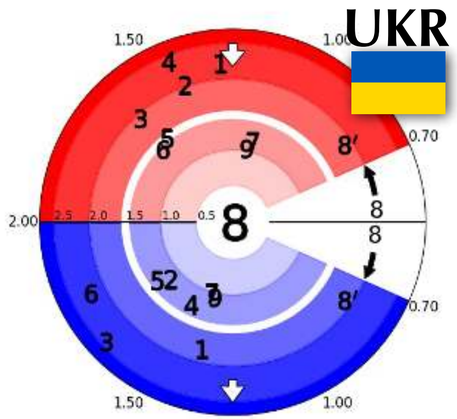


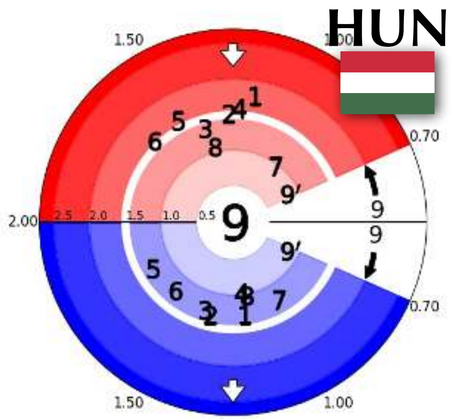




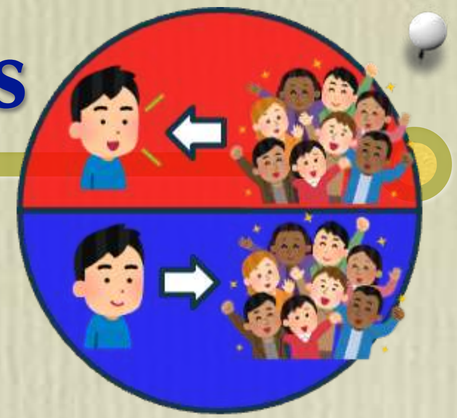




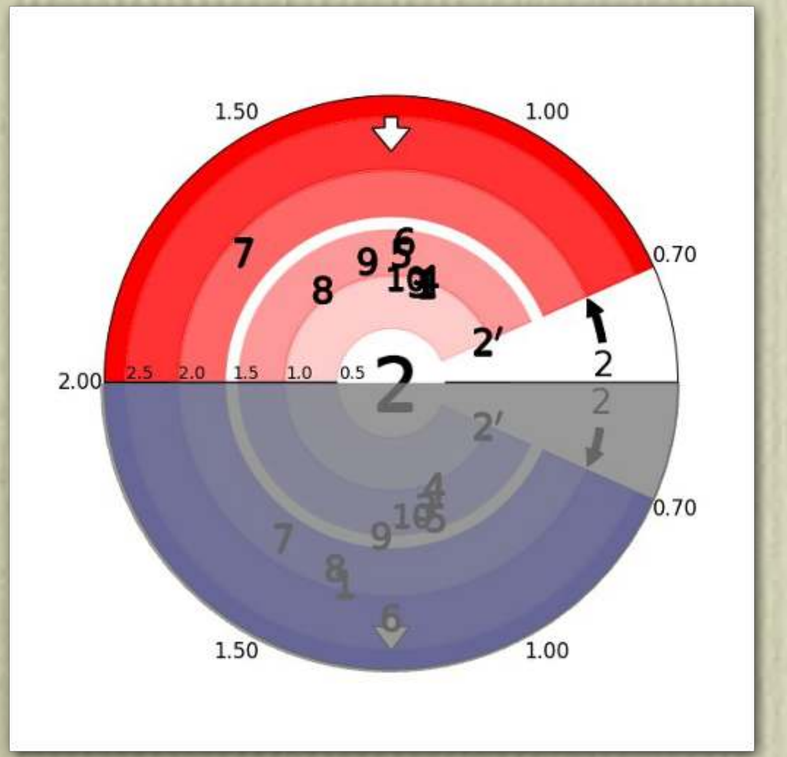




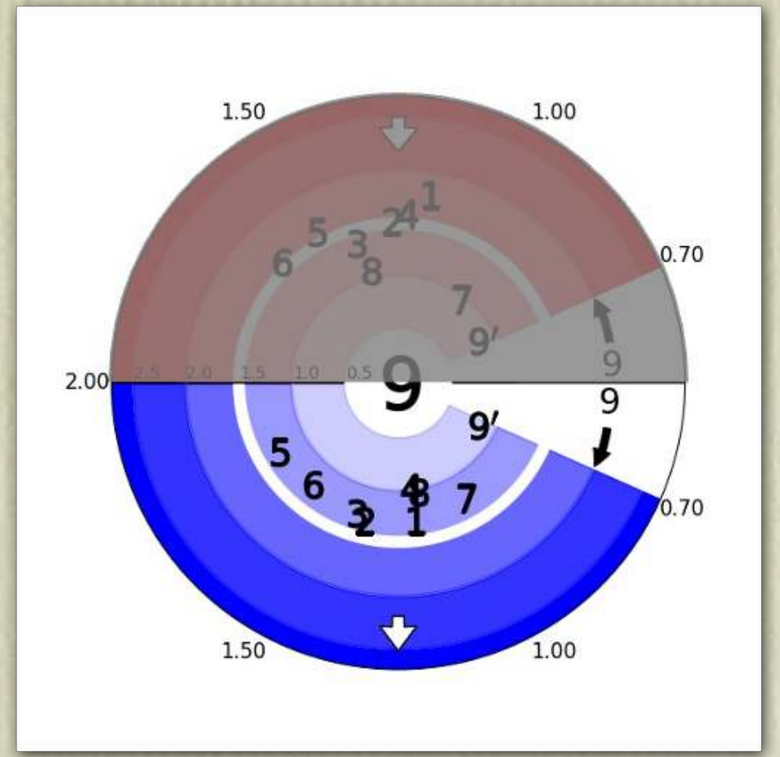
Salient patterns found in the 28 C-charts



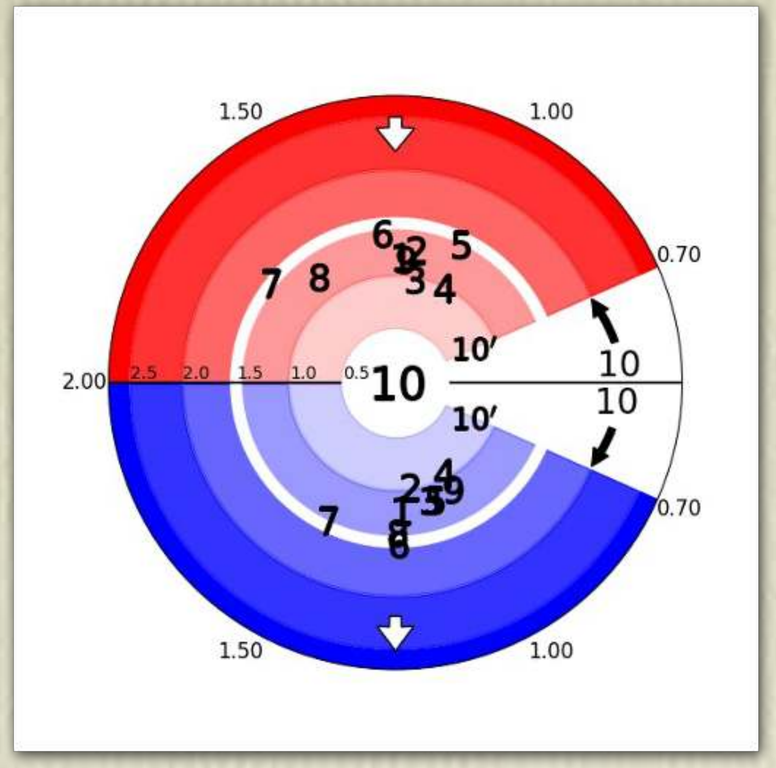
Patterns (a) to (c)



(a) Intelligible speaker

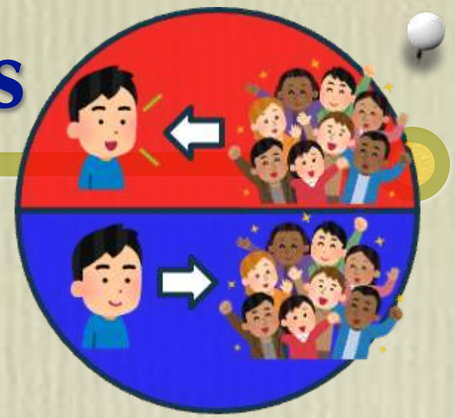


(b) Fluent listener

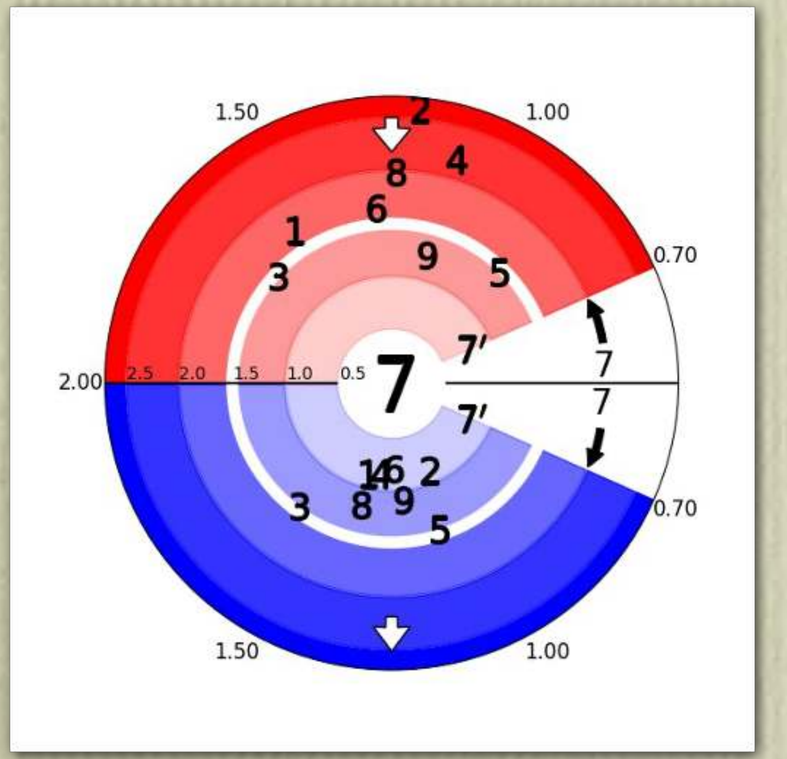


(c) Ideal learner = (a)+(b)

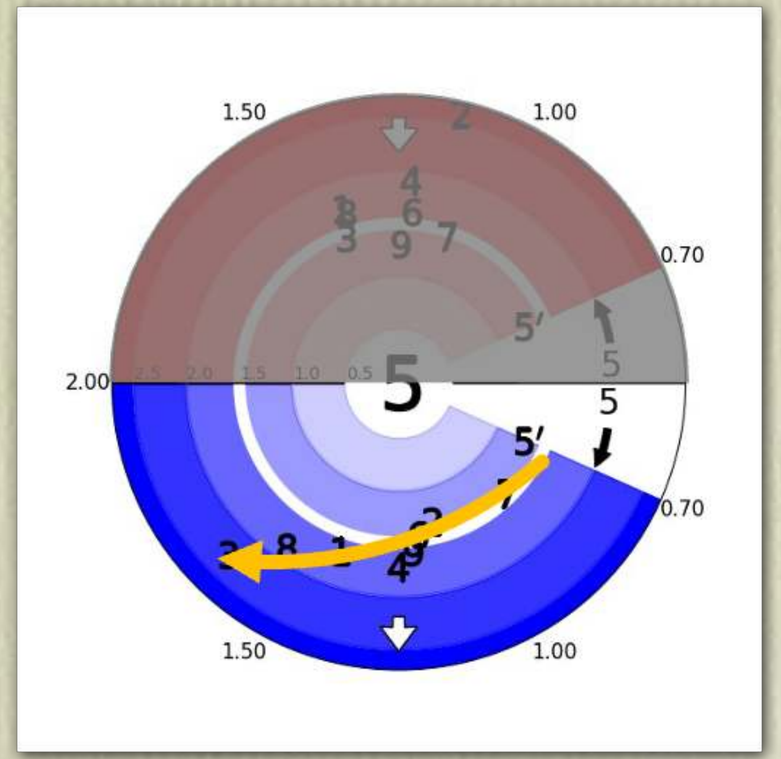
Salient patterns found in the 28 C-charts



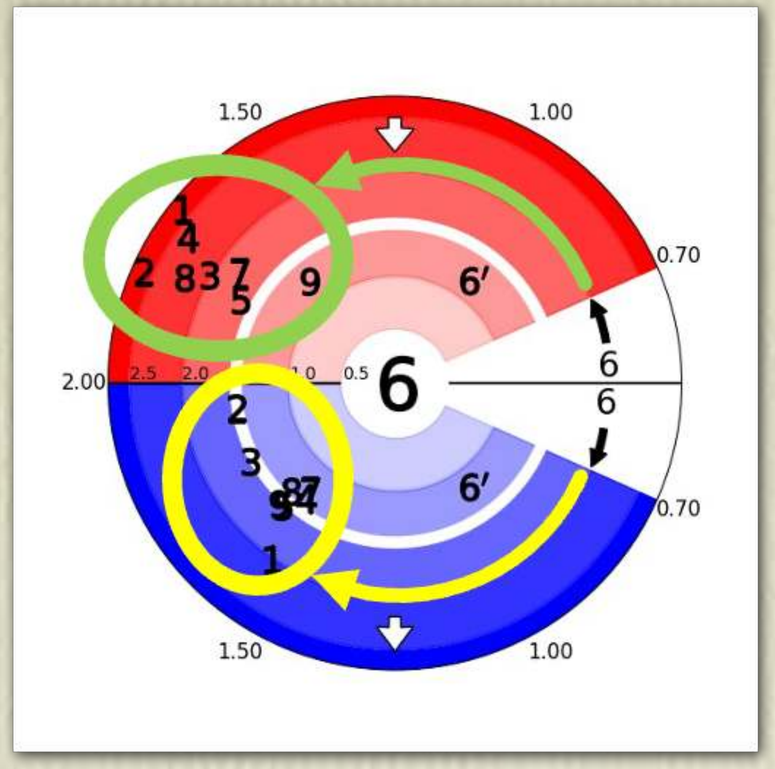
Patterns (d) to (f)



(d) Fluent listener but unintelligible speaker



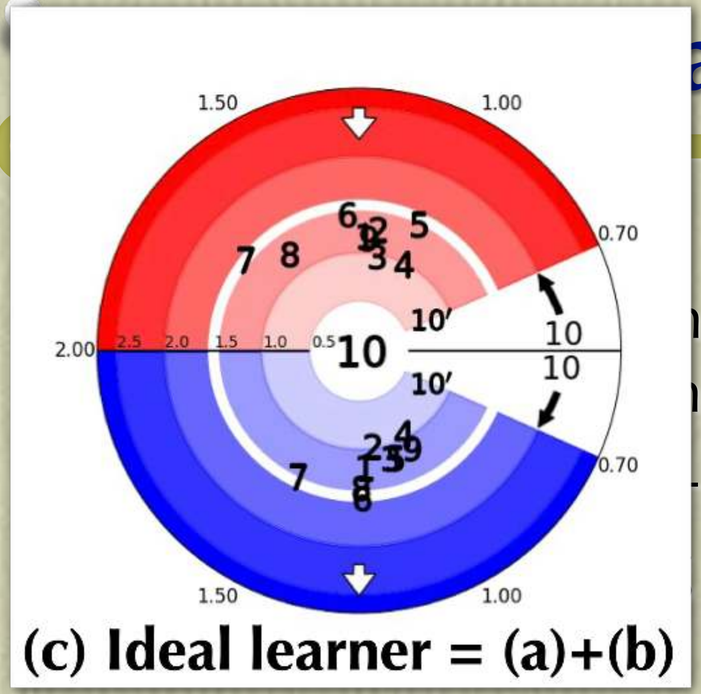
(e) Dependent listening



(f) Speaker with a very unique accent

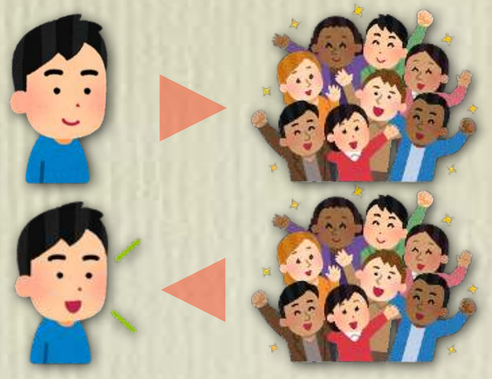
al shadowing among WE speakers

the mutual shadowing experiment



(c) Ideal learner = (a)+(b)

hese classes at UTokyo and divided into 3 groups with good care paid
nguage background of each group's participants.
-sec length were selected from Eiken Grade-2 Test (ARI = 6.2 to 7.0).
experiment was conducted for each group.
everybody, and everybody was shadowed by everybody.



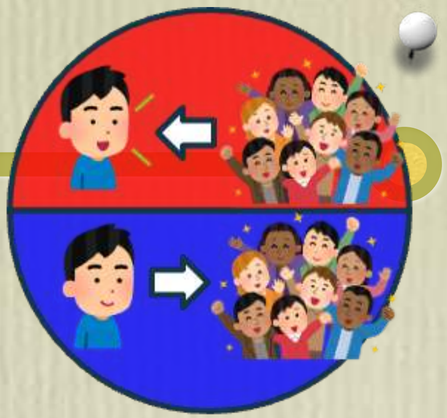
ID	lang.	group A family	C-chart
1	CHN	ST	e
2	CHN	ST	a,e
3	JPN	TU	a,e
4	KOR	TU	a,e
5	KOR	TU	a,e
6	FRA	IE (IT)	d,e
7	ITA	IE (IT)	c̄,f
8*	HIN	IE (II)	d,f
9	SRB	IE (SL)	c,e
10	HUN	UR (FU)	c,e

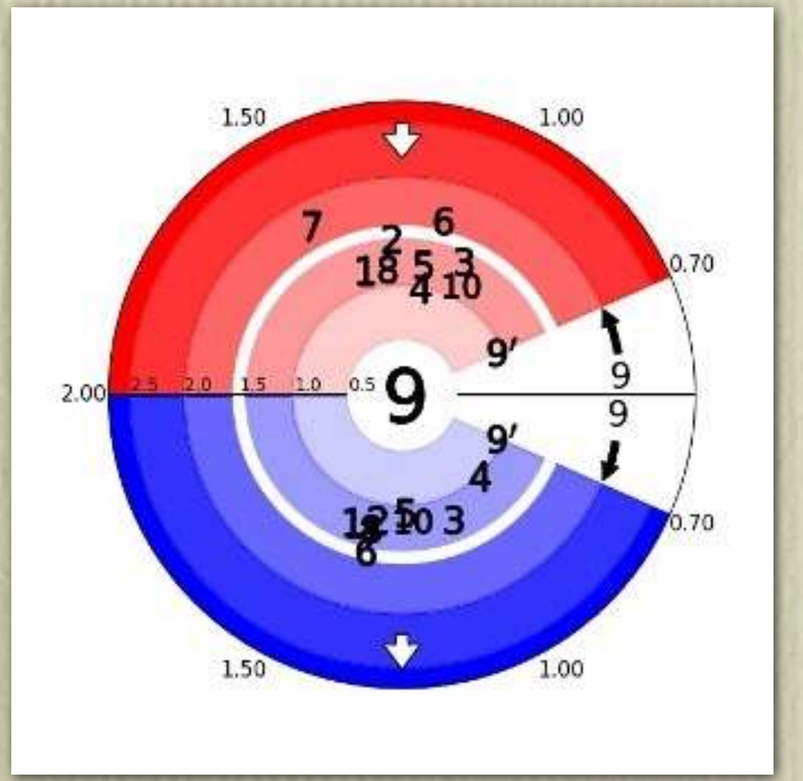
ID	lang.	group B family	C-chart
1	CHN	ST	b̄,e,f
2	CHN	ST	d̄
3	JPN	TU	ā,f
4	KOR	TU	b̄,e
5	FRA	IE (IT)	c̄,e
6	ITA	IE (IT)	c̄,e
7*	HIN	IE (II)	d
8	UKR	IE (SL)	c̄,f
9*	MAL	DR	d

ID	lang.	group C family	C-chart
1	CHN	ST	c̄,e
2	CHN	ST	d̄,e
3	JPN	TU	b̄,e,f
4	KOR	TU	b̄,e
5	FRA	IE (IT)	b̄,f
6	SPN	IE (IT)	d,f
7*	HIN	IE (II)	b,e
8	UKR	IE (SL)	f
9	HUN	UR (FU)	c

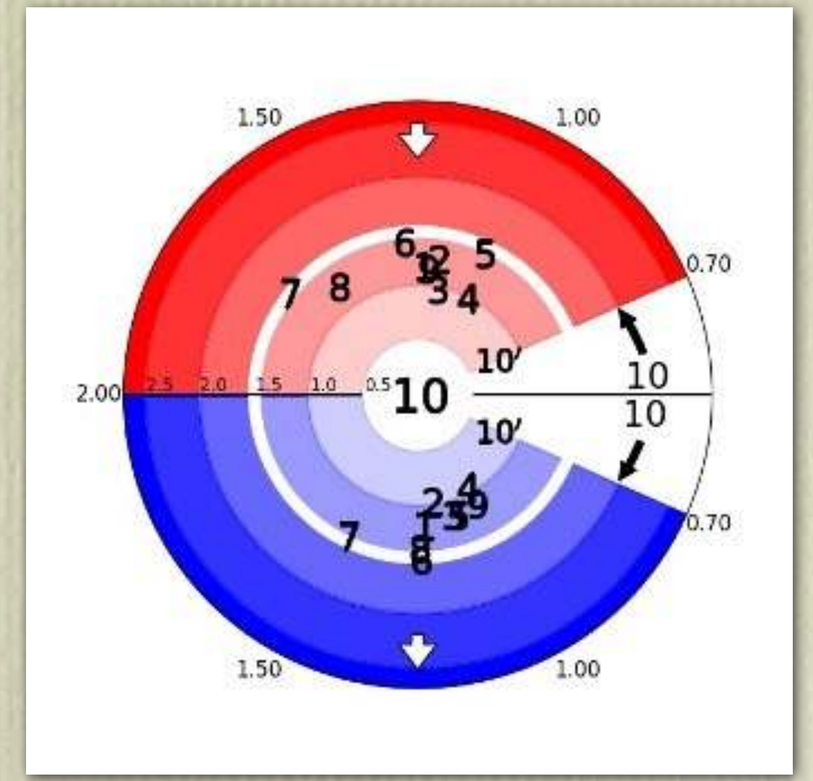
languages		lang. families	
CHN	Chinese	ST	Sino-Tibetan
JPN	Japanese	TU	Trans-Eurasian
KOR	Korean	IE	Indo-European
FRA	French	UR	Uralic
ITA	Italian	DR	Dravidian
SPN	Spanish		
HIN	Hindi	lang. sub-families	
SRB	Serbian	IT	Italic
UKR	Ukrainian	II	Indo-Iranian
HUN	Hungarian	SL	Slavic
MAL	Malayāḷam	FU	Finno-Ugric

Super learners?

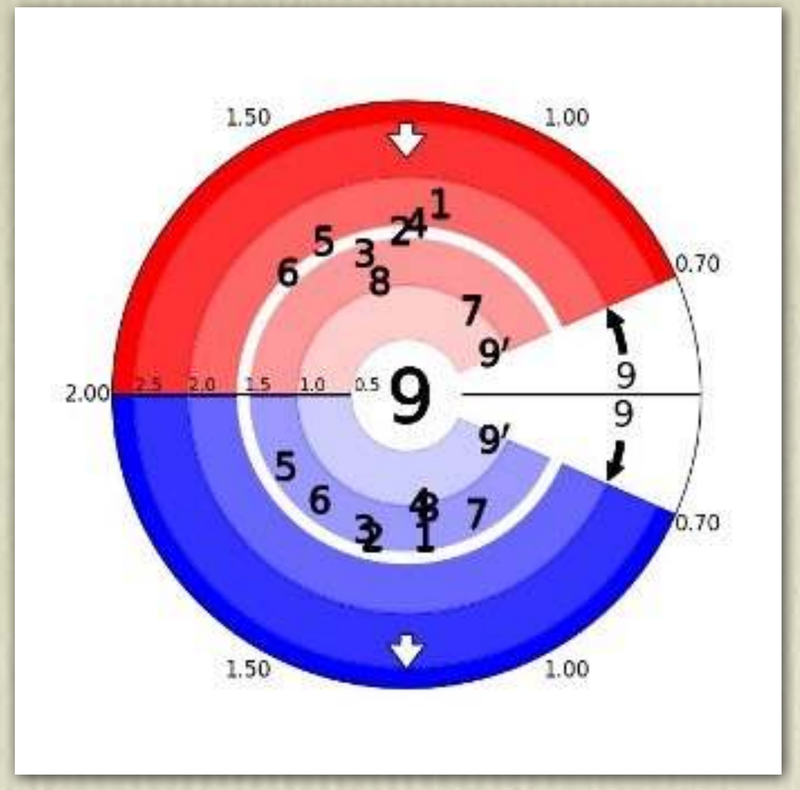




SRB in group A



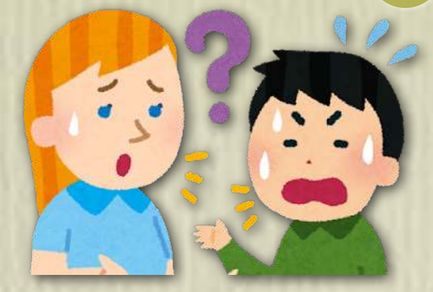
HUN in group A

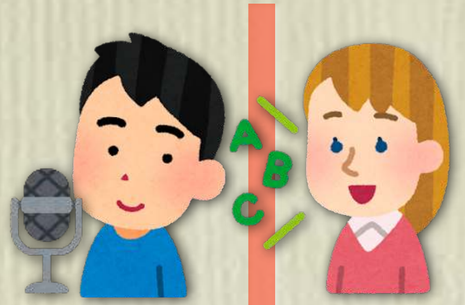


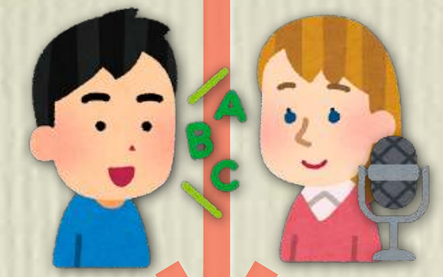
HUN in group C

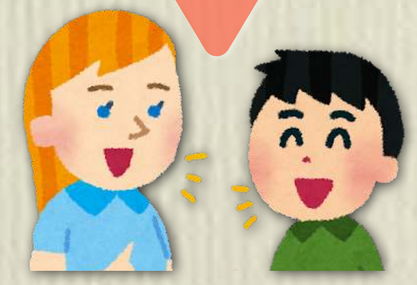
Outline of this talk

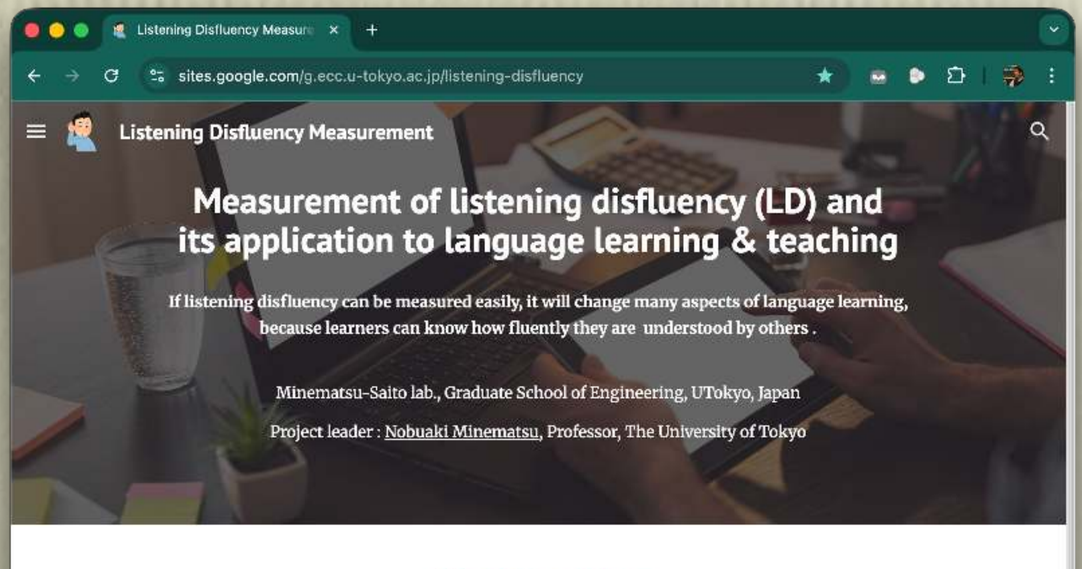
- Why listening disfluency (LD)?
- How to measure LD with a microphone?
- Measurement and analysis of learners' LD
- Measurement and analysis of raters' LD
- Prediction of raters' LD
- Conclusions







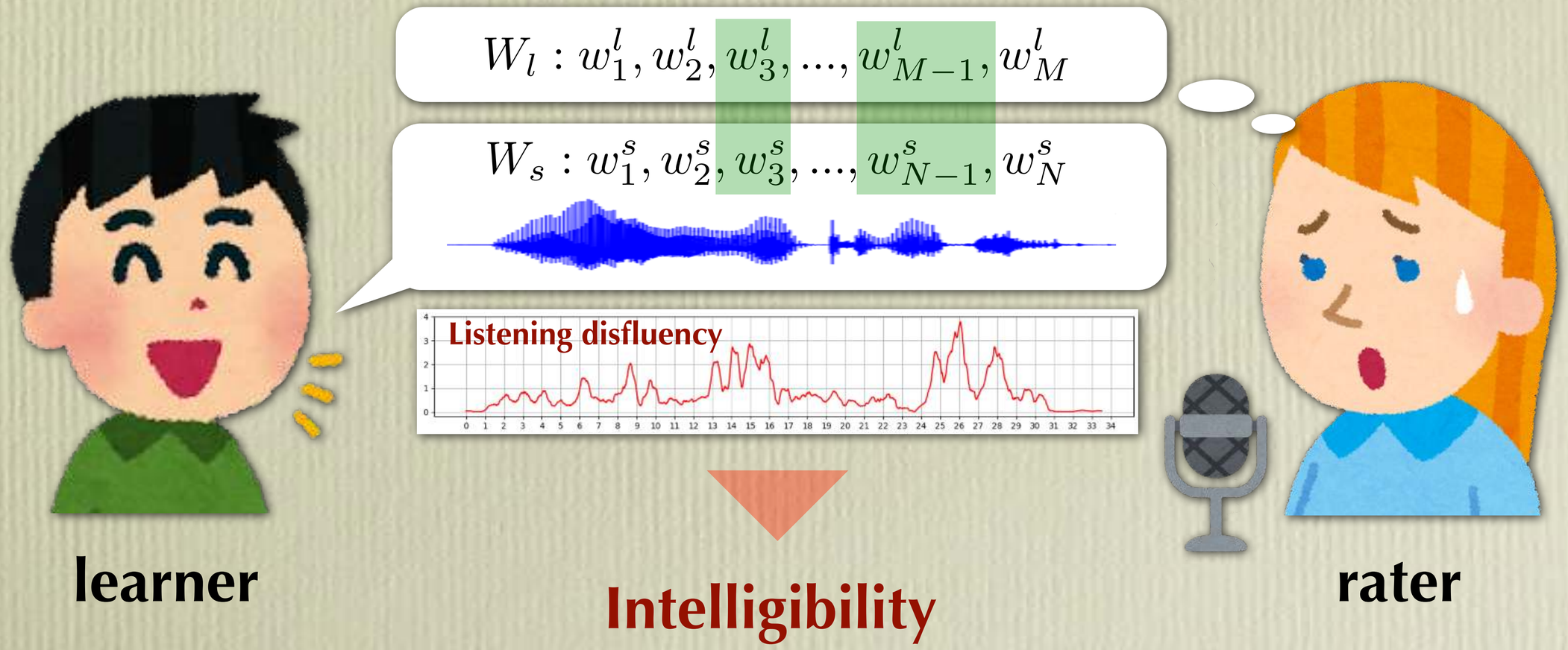




Assessment of speaking based on listening [Inoue+'18]

When listening, where in a given speech does LD take place?

- Listening is mental activity and does not present any acoustic events.

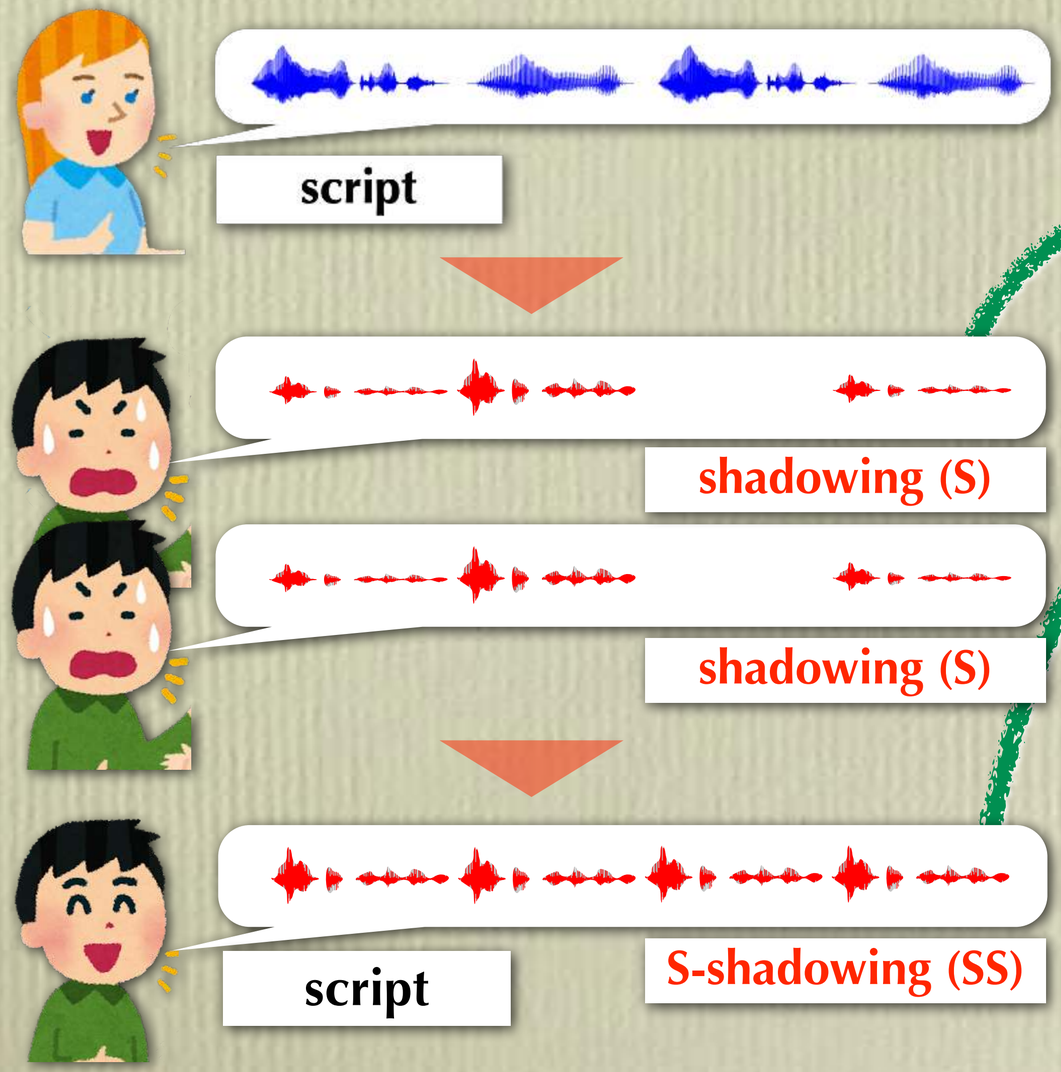


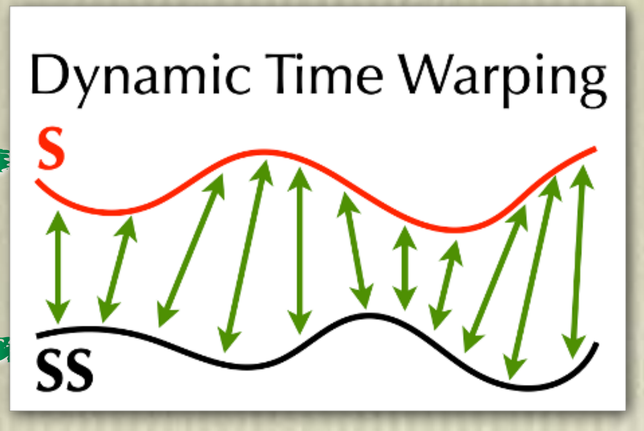
Intelligibility

Inoue, Y., et al. (2018) A Study of Objective Measurement of Comprehensibility through Native Speakers' Shadowing of Learners' Utterances. Proc. Interspeech 2018, 1651-1655

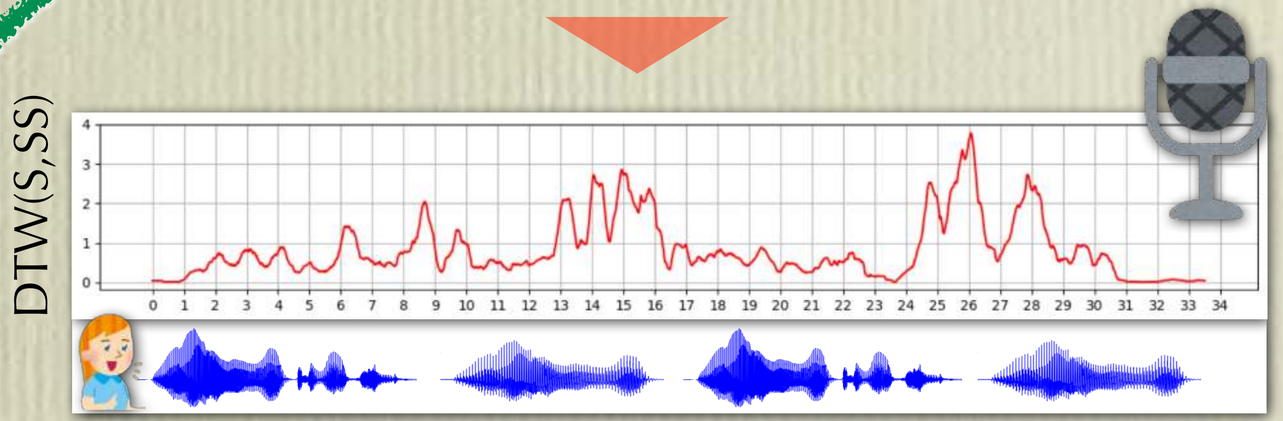
Measurement of listening in class

Dictation test by speaking using a mouth [Inoue+'18, Zhu+'20]





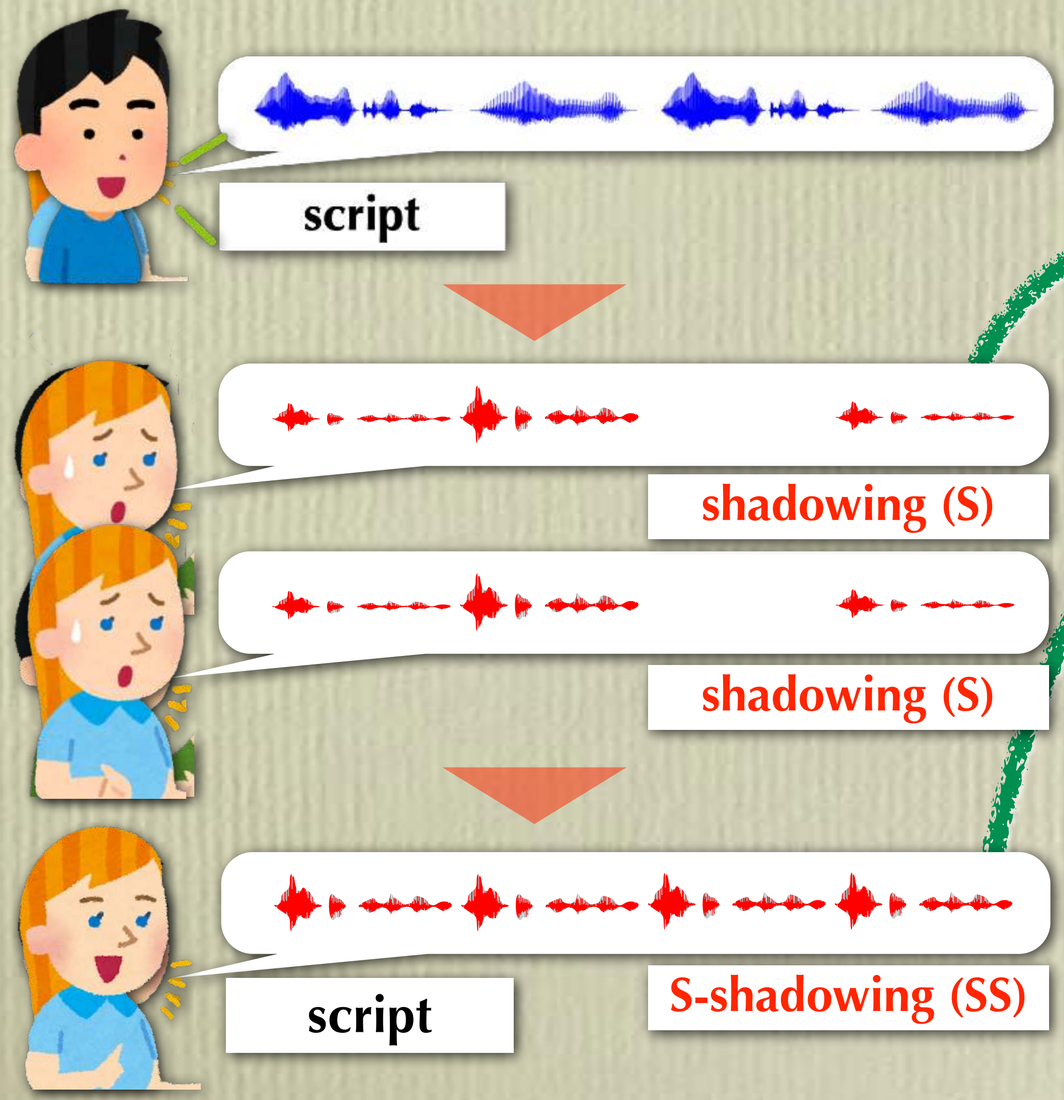
PPG-DTW
(**P**honetic
Posterior**G**ram)

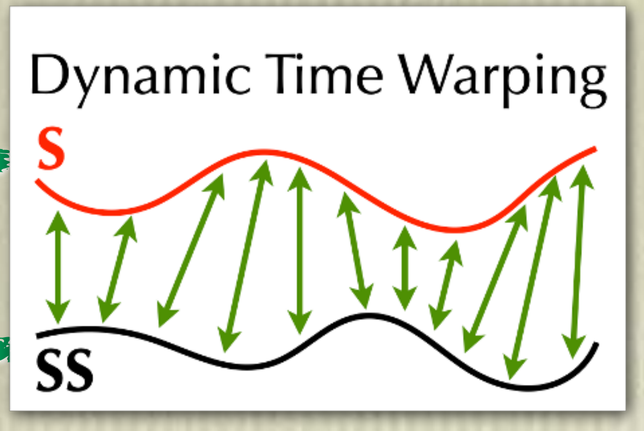


Shadowing disfluency = listening disfluency

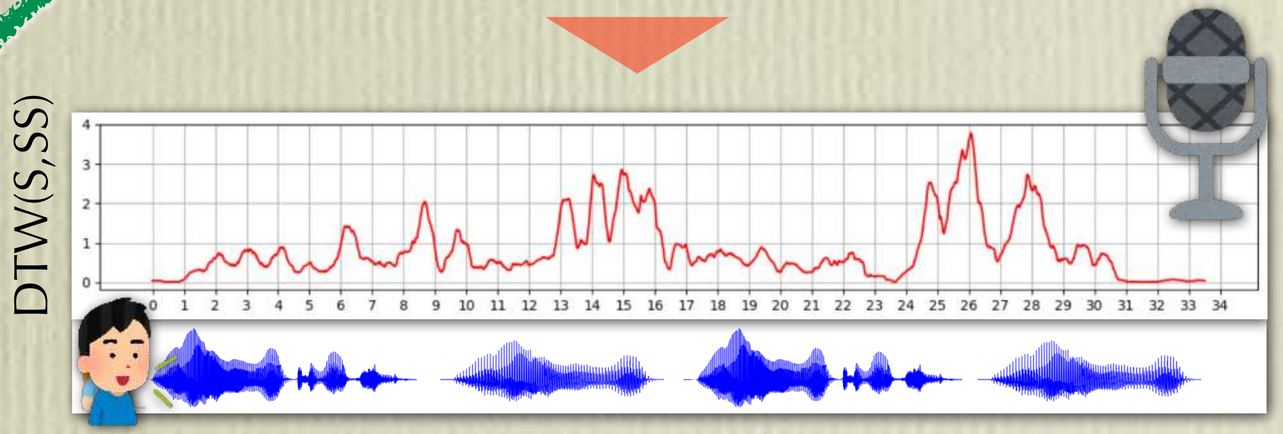
Measurement of listening in class

Dictation test by speaking using a mouth [Inoue+'18, Zhu+'20]





PPG-DTW
(**P**honetic
Posterior**G**ram)



Shadowing disfluency = listening disfluency

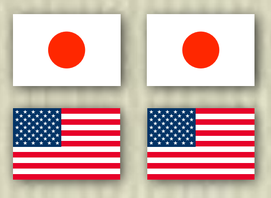
Listening behaviors of raters with different profiles [Zhu+'21]

Speakers and their reading-alouds

- 12 Japanese learners of English selected out of 30 to cover a wide range of fluency
- 2 native speakers (AE and BE)
- The 14 speakers read aloud about 30-sec-long passages extracted from textbooks.

7 raters with different language profiles

- 2 Japanese with advanced proficiency of English **(J1, J2)**
- 2 Americans who do not speak Japanese **(N1, N2)**
- 2 Chinese and 1 Vietnamese with advanced proficiency of English **(NN1, NN2, NN3)**



Task

- The raters shadow and script-shadow the 14 passages from the 14 speakers.
  - Their shadowing utterances are dictated by those whose L1 is the same as L1 of the raters.
  - Intelligibility of the 14 passages to the 7 raters is quantified using PPG-DTW(S,SS).

C. Zhu, et al, "Multi-Granularity Annotation of Instantaneous Intelligibility of Learners' Utterances Based on Shadowing Techniques," 2021 Proc. ASRU, 2021, pp. 1071-1078

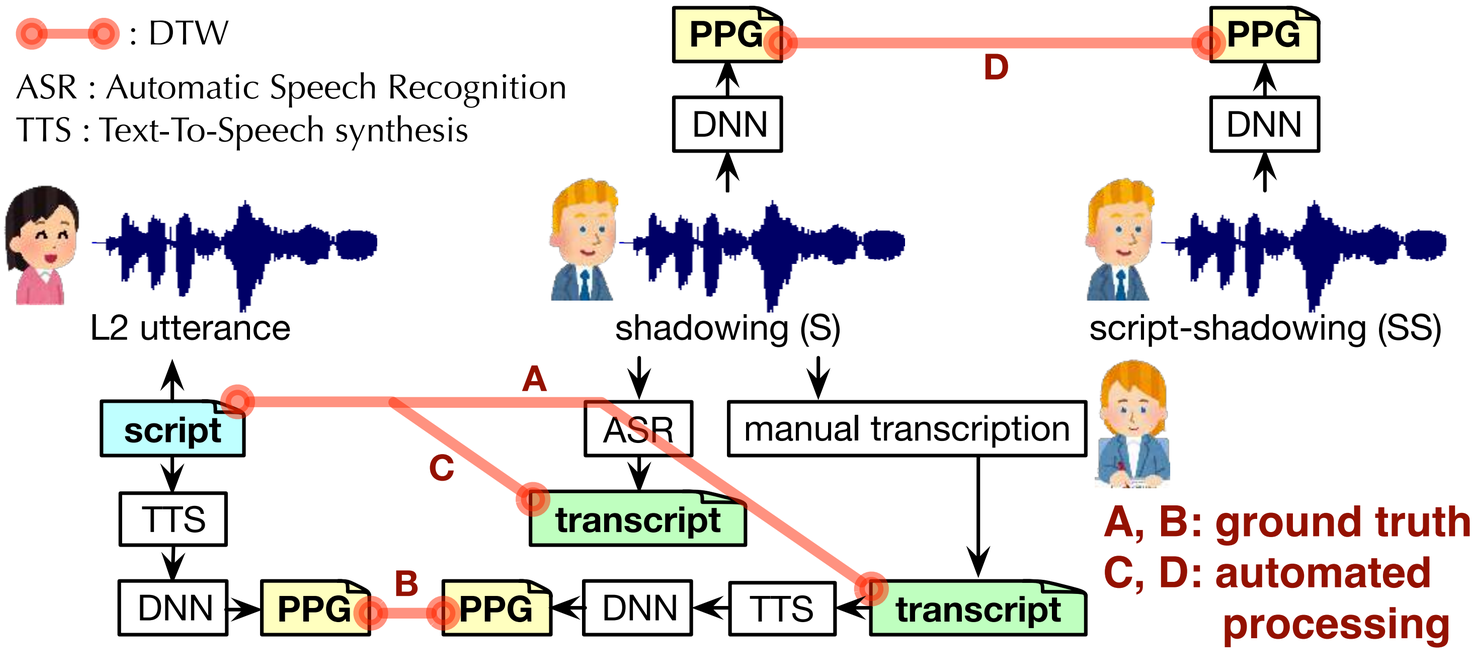
Listening behaviors of raters with different profiles [Zhu+'21]

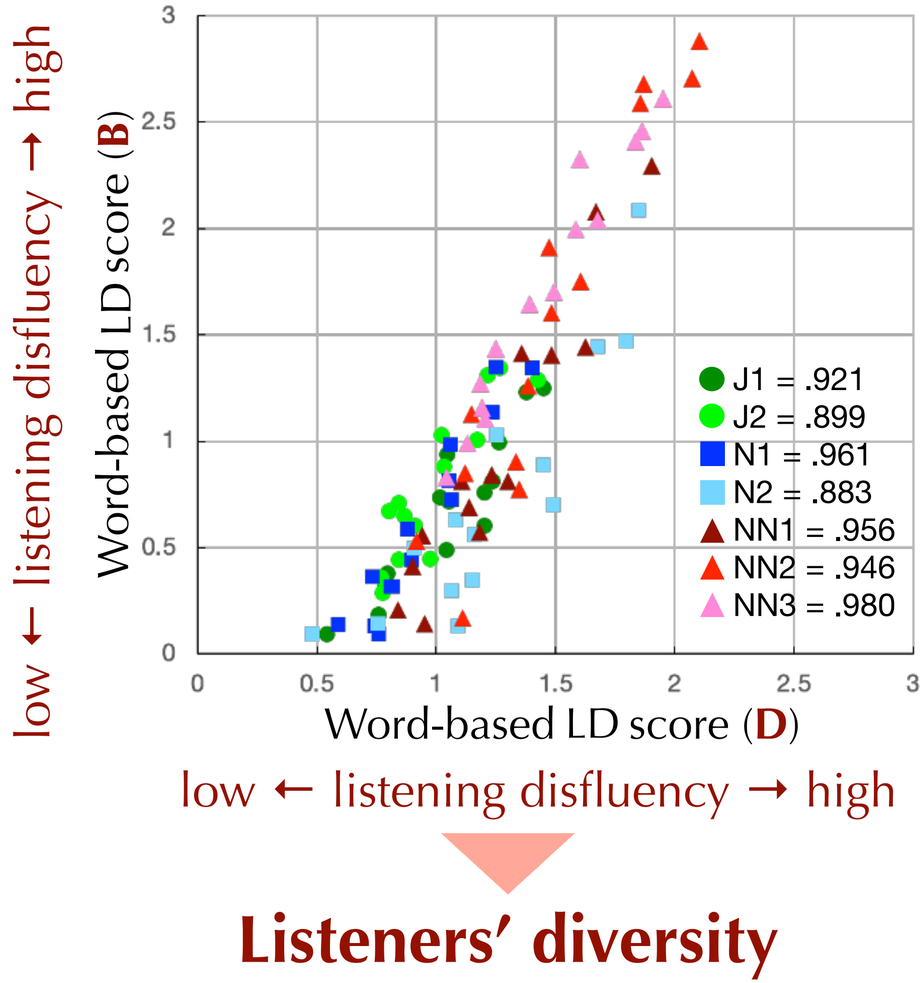
Results of raters' shadowing experiments

- Japanese (J1, J2), American (N1, N2), and non-native shadowers (NN1, NN2, NN3)

Corr.	J1	J2	N1	N2	NN1	NN2	NN3	mean
C–A	0.874	0.690	0.889	0.854	0.904	0.910	0.847	0.853
D–A	0.869	0.898	0.869	0.860	0.937	0.916	0.901	0.893
D–B	**0.921**	**0.899**	**0.961**	**0.883**	**0.956**	**0.946**	**0.980**	**0.935**

Correlations of D–A are all negative, but the signs are removed.

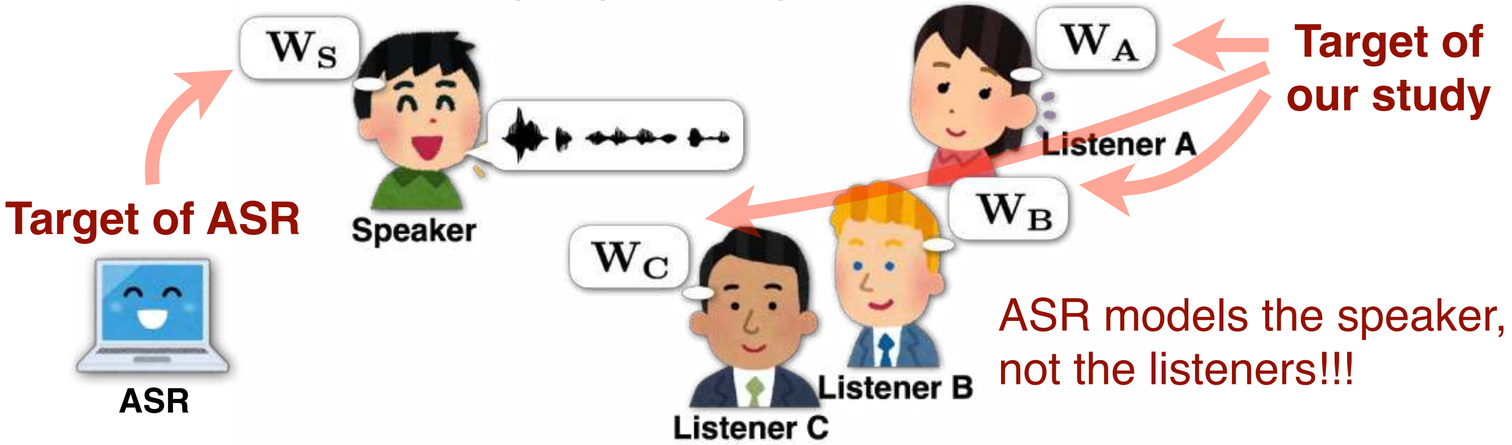




L2 speech corpus with (un)intelligibility annotation

L2 speech corpus with intelligibility annotation

- **Different listeners show different listening performances.**
  - Depend on their language backgrounds.



- **Learners (speakers) and their reading-alouds**
  - Approx. 8,700 original passages from approx. 250 learners
    - CEFR = A1, A2, and B1. Recording was made online.
  - Preprocessing and selection conducted on the passages
    - Segmentation into about 30-sec speech segments
    - Removal of segments including long pauses, noise, or rare words
    - Text-based clustering to tens of clusters for topic diversity
    - 753 oral passages with various topics were selected.

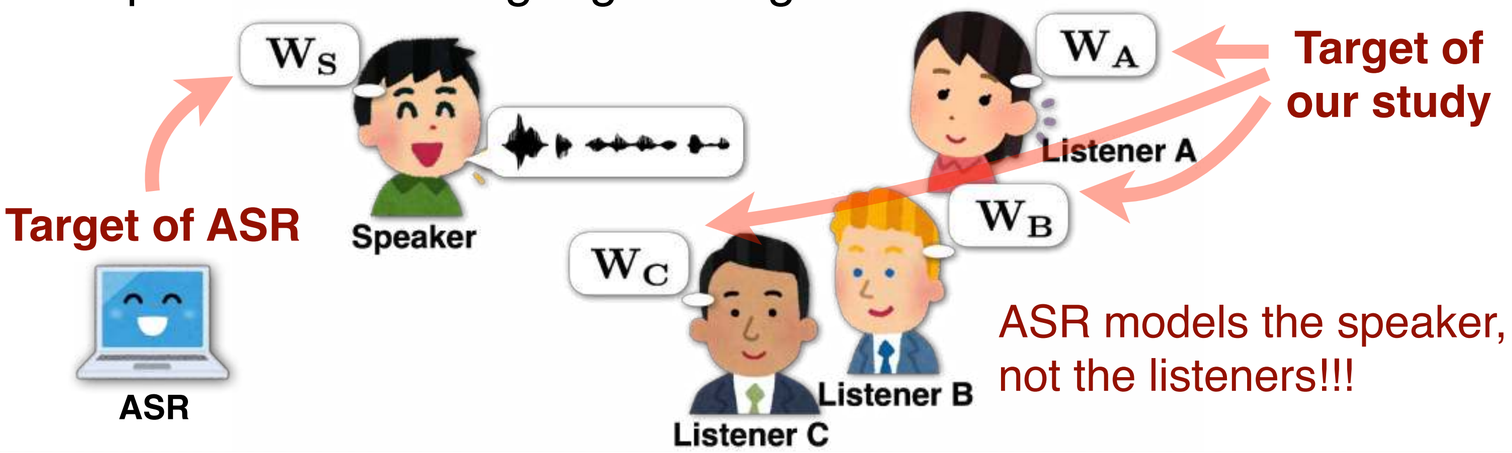
- Non-native ASR is one of the hot topics of ASR research.
- State-of-the-art ASR is sometimes too good to be used in class.
- NN-to-NN communication inevitably involves miscommunication.
- Human-like misrecognition systems are needed.
  = Listener simulator

C. Zhu, et al."Automatic Prediction of Intelligibility of Words and Phonemes Produced Orally by Japanese Learners of English," 2022 IEEE Spoken Language Technology Workshop (SLT), 2023, pp. 1029-1036

L2 speech corpus with (un)intelligibility annotation

L2 speech corpus with intelligibility annotation

- **Different listeners show different listening performances.**
  - Depend on their language backgrounds.



- **Three shadowers with different language backgrounds**
  - **A** is a doctoral student of applied linguistics.
    - L1 is Japanese. She had stayed in UK and Canada for 7 years.
  - **B** is a bachelor who is an intermediate learner of Japanese.
    - L1 is US English. He planned to teach English in Japan.
  - **C** is a master student of musicology who does not speak J.
    - L1 is US English. Collection of his data will be finished very soon.

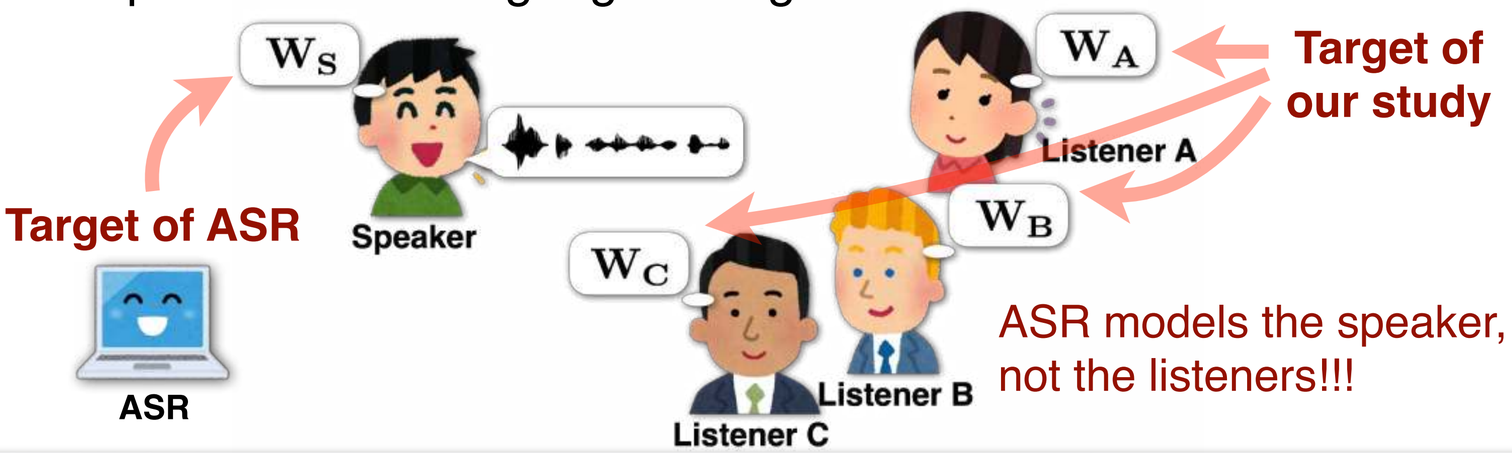
- Non-native ASR is one of the hot topics of ASR research.
- State-of-the-art ASR is sometimes too good to be used in class.
- NN-to-NN communication inevitably involves miscommunication.
- Human-like misrecognition systems are needed.
  = Listener simulator

C. Zhu, et al."Automatic Prediction of Intelligibility of Words and Phonemes Produced Orally by Japanese Learners of English," 2022 IEEE Spoken Language Technology Workshop (SLT), 2023, pp. 1029-1036

L2 speech corpus with (un)intelligibility annotation

L2 speech corpus with intelligibility annotation

- **Different listeners show different listening performances.**
  - Depend on their language backgrounds.



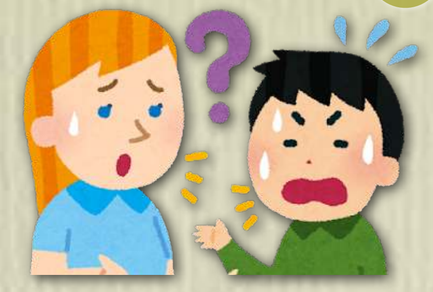
- **Recording of three shadowings and DTW-based analysis**
  - Shadow (**S1**), shadow (**S2**), and script-shadow (**SS**)
  - Online recording was made with the same type of headset.
  - PPG-DTW(**S1**,**SS**) and PPG-DTW(**S2**,**SS**)
    - Sn and SS were converted into their PPG with WSJ-Kaldi.
  - PPG-DTW(**Sn**,**SS**) was mapped on learners' utterances.
  - Thresholding to define (un)intelligible word/phoneme segments in the learners' utterances

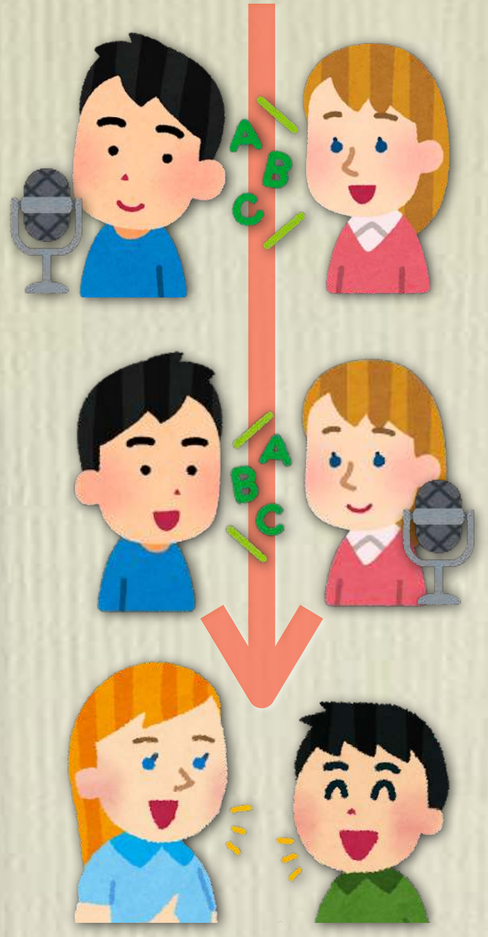
- Non-native ASR is one of the hot topics of ASR research.
- State-of-the-art ASR is sometimes too good to be used in class.
- NN-to-NN communication inevitably involves miscommunication.
- Human-like misrecognition systems are needed.
  = Listener simulator

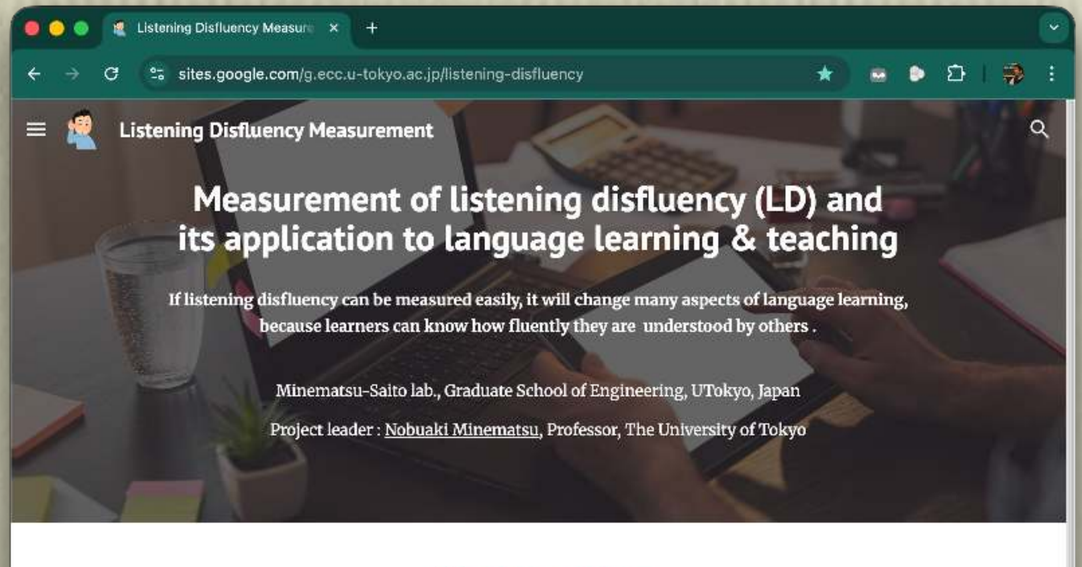
C. Zhu, et al."Automatic Prediction of Intelligibility of Words and Phonemes Produced Orally by Japanese Learners of English," 2022 IEEE Spoken Language Technology Workshop (SLT), 2023, pp. 1029-1036

Outline of this talk

- Why listening disfluency (LD)?
- How to measure LD with a microphone?
- Measurement and analysis of learners' LD
- Measurement and analysis of raters' LD
- Prediction of raters' LD
- Conclusions







Simulating native shadowers' performance

Applying voice conversion techniques to simulate native shadowers

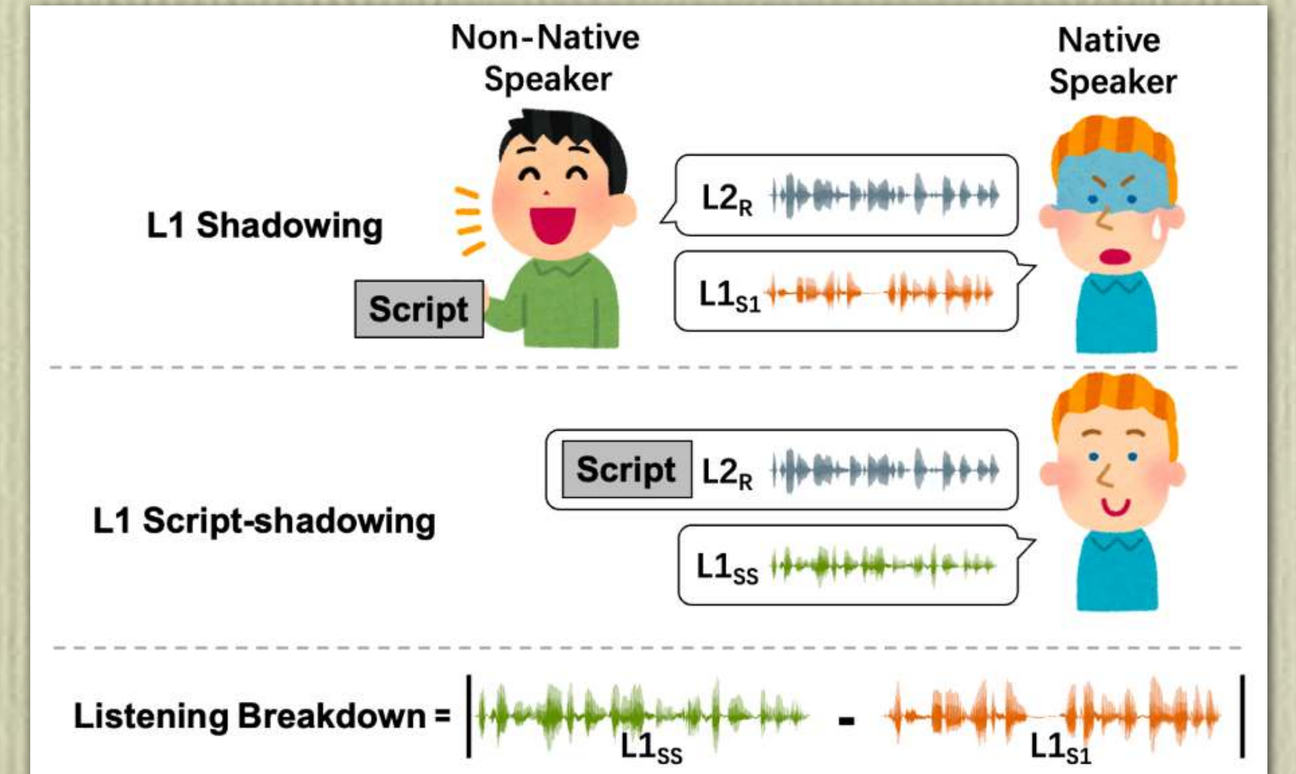


Fig. 1: Native speaker shadowing for non-native speech assessment. $L1_{S1}$ represents a native speaker's initial shadowing, while $L1_{SS}$ denotes his/her script-shadowing, which is the most fluent shadowing.

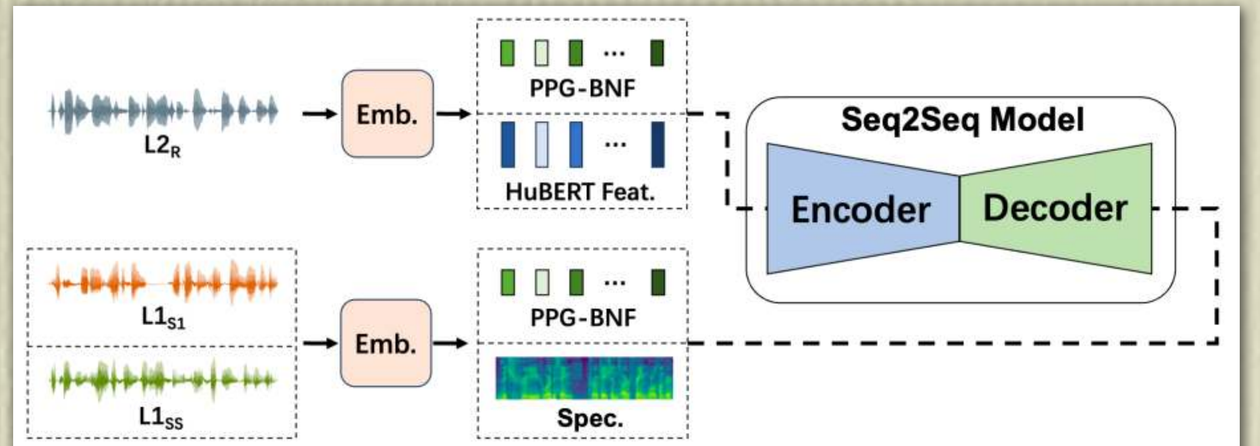


Fig. 2: Overview of the proposed L1-shadowing-L2 system: After embedding the source and target features with respective encoders, the model directly transforms the source input $L2_R$ into target outputs $L1_{S1}$ or $L1_{SS}$. Unlike previous methods using PPG-BNF and mel-spectrogram features, our approach leverages self-supervised feature, specifically, HuBERT feature, in the source phase to capture the general characteristics of L2 utterances.

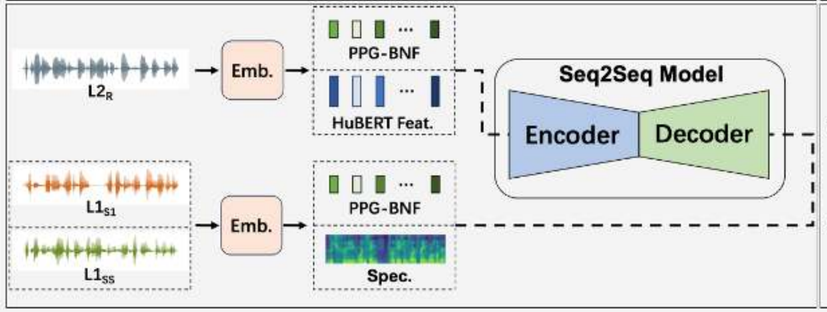
Demo

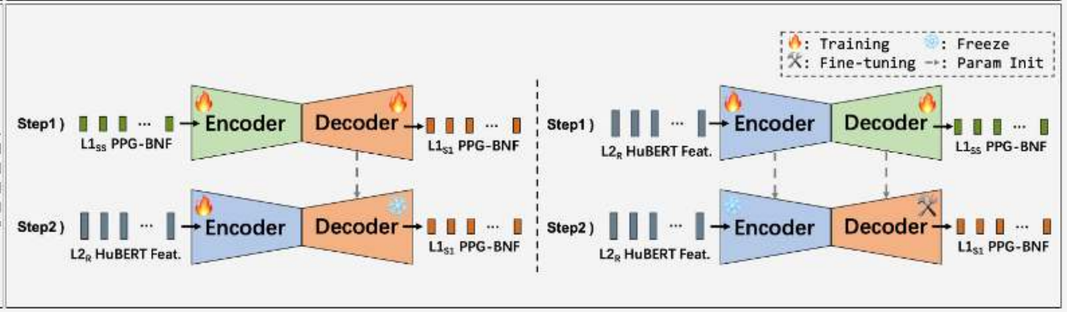
ICASSP_2025_Virtual_Shado

secondtonumb.github.io/publication_demo/ICASSP_2025/index.html

Proposed Methods

Native Speaker Shadowing System with Latent Speech Representations	2 Step Training Methods





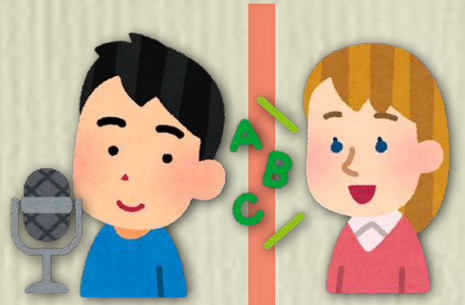
Audio Samples (Click ID to display reference text)

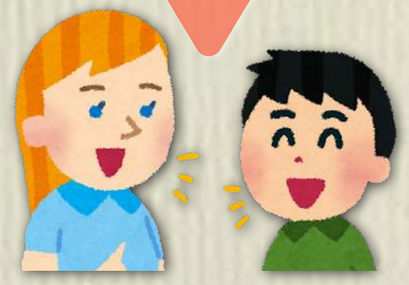
0668-0004

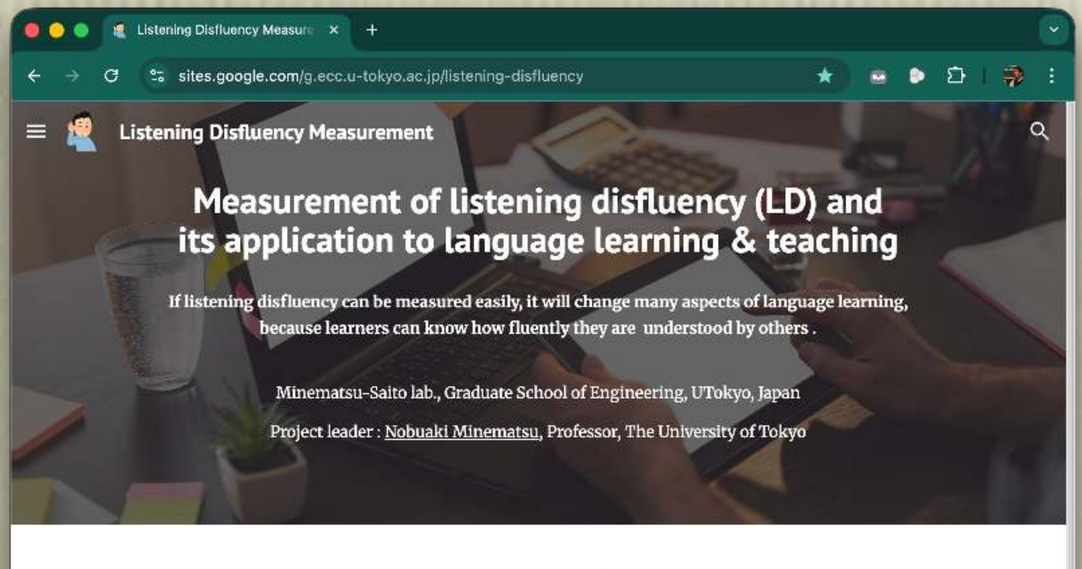
Source		Target								Reference	
$L2_R$		Mel.				PPG-BNF				$L1_{S1}$	$L1_{SS}$
		$L1_{S1}$	$L1_{SS}$	$L1_{S1}$ Enc.FT	$L1_{S1}$ Dec.FT	$L1_{S1}$	$L1_{SS}$	$L1_{S1}$ Enc.FT	$L1_{S1}$ Dec.Ft		
▶	PPG-BNF	▶	▶			▶	▶			▶	▶
	HuBERT	▶	▶	▶	▶	▶	▶	▶	▶		

Outline of this talk

- Why listening disfluency (LD)?
- How to measure LD with a microphone?
- Measurement and analysis of learners' LD
- Measurement and analysis of raters' LD
- Prediction of raters' LD
- Conclusions







Assessment of listening [Inoue+'18]

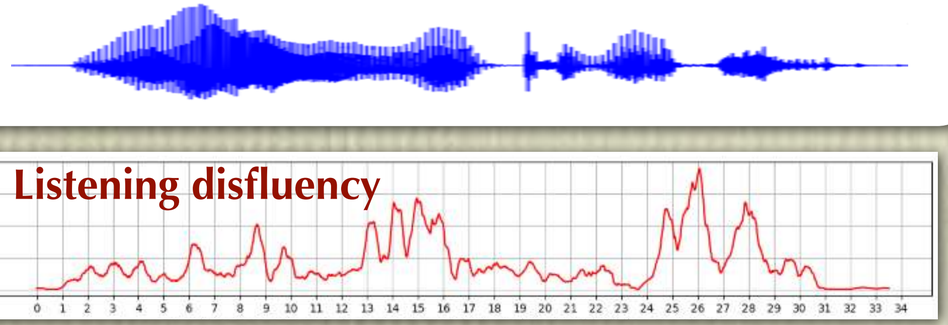
When listening, where in a given speech does LD take place?

- Listening is mental activity and does not present any acoustic events.

native

$$W_l : w_1^l, w_2^l, w_3^l, \ldots, w_{M-1}^l, w_M^l$$

$$W_s : w_1^s, w_2^s, w_3^s, \ldots, w_{N-1}^s, w_N^s$$



https://www.artinis.com/nirs-devices

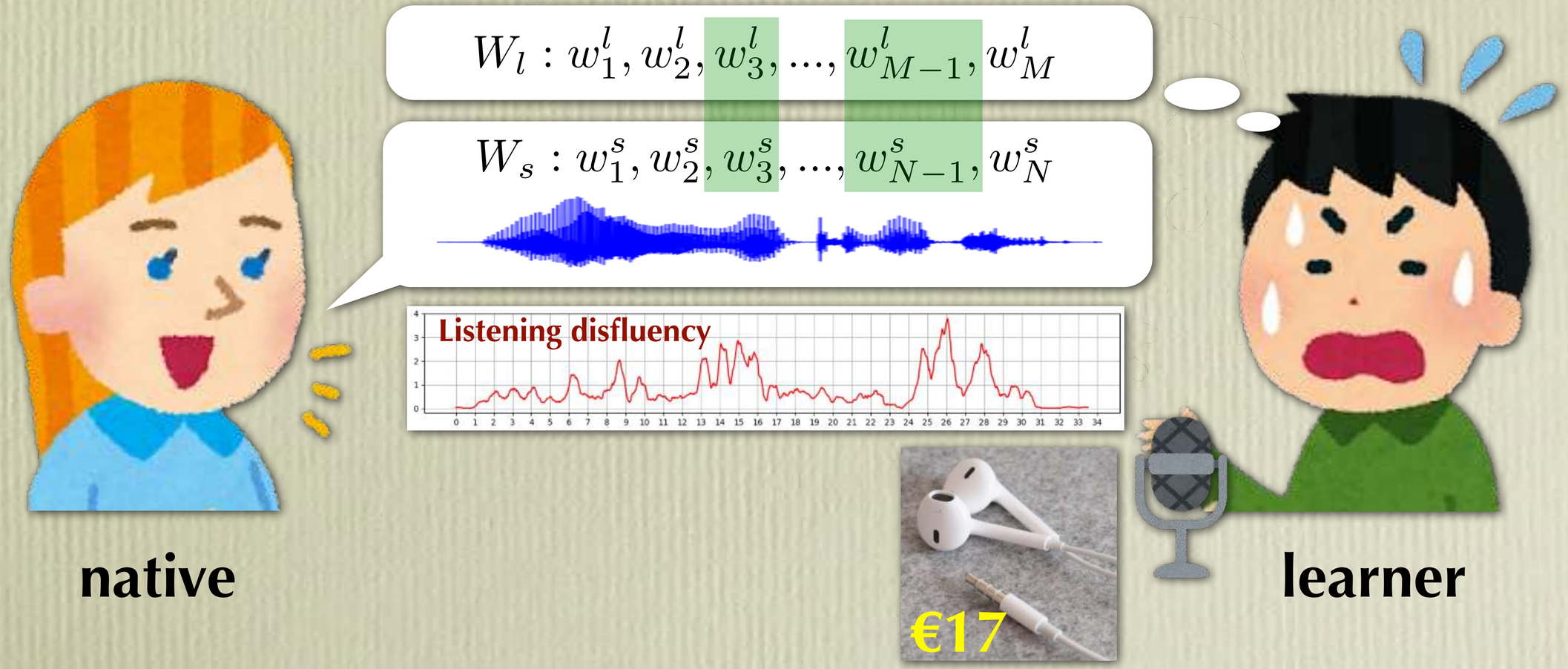
learner

Inoue, Y., et al. (2018) A Study of Objective Measurement of Comprehensibility through Native Speakers' Shadowing of Learners' Utterances. Proc. Interspeech 2018, 1651-1655

Assessment of listening [Inoue+'18]

When listening, where in a given speech does LD take place?

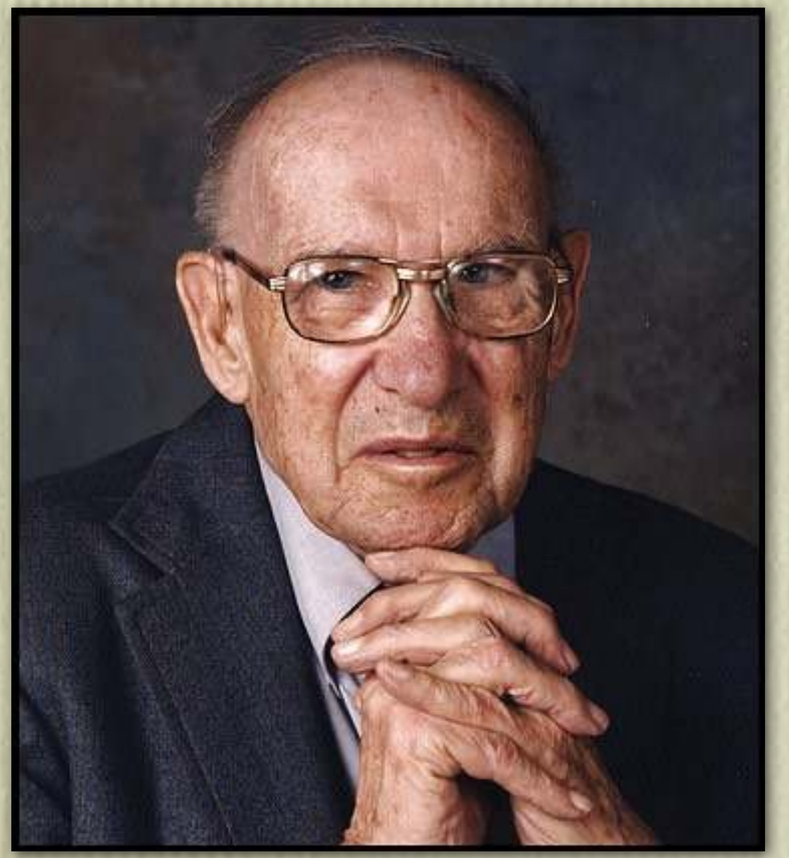
- Listening is mental activity and does not present any acoustic events.



Inoue, Y., et al. (2018) A Study of Objective Measurement of Comprehensibility through Native Speakers' Shadowing of Learners' Utterances. Proc. Interspeech 2018, 1651-1655

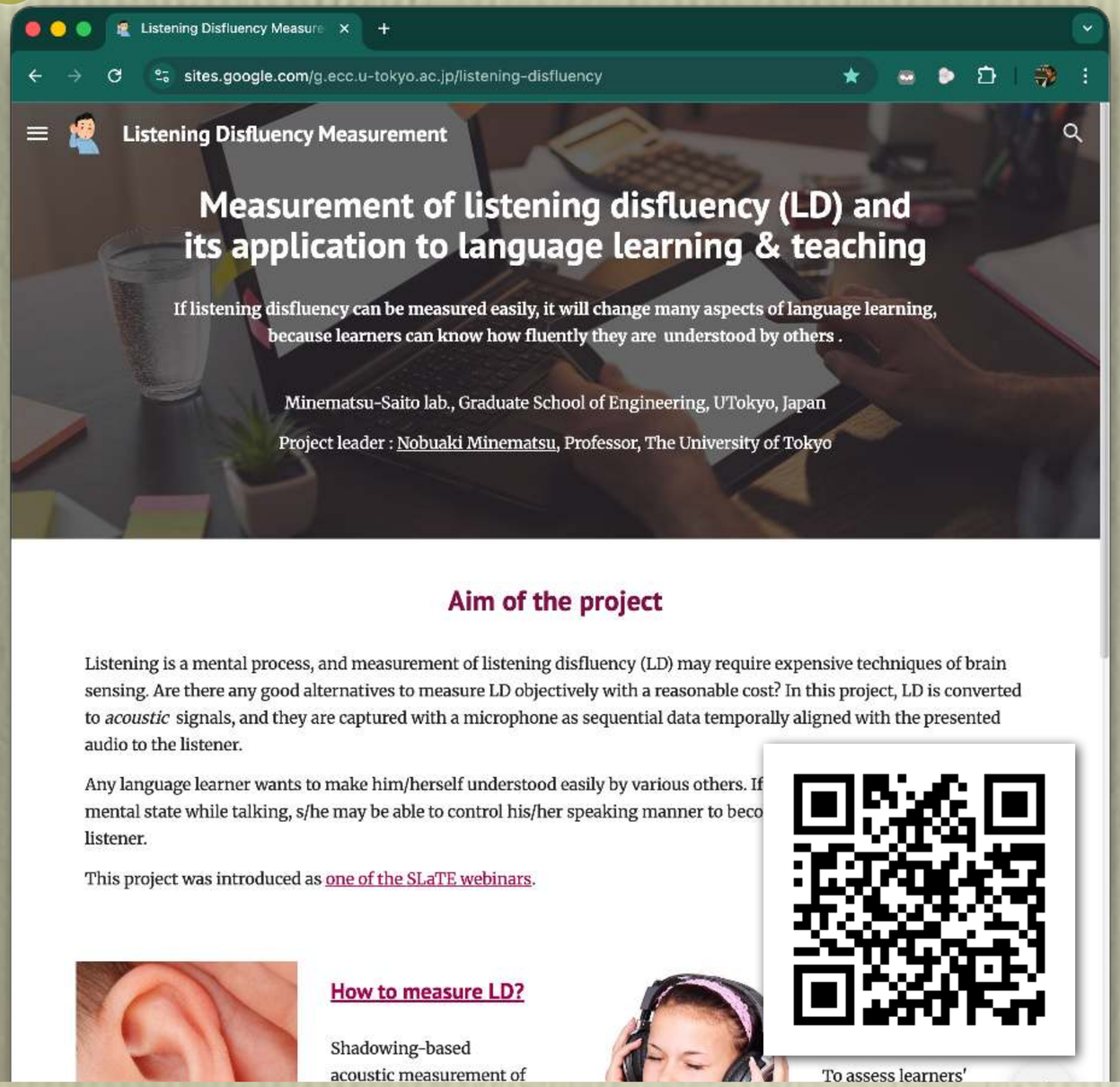
A message to all of you.

Peter F. Drucker, the father of modern management



"What gets measured gets improved."

Our project and our course



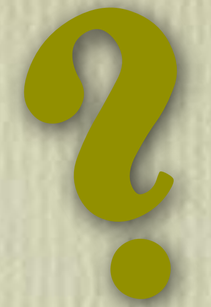
聞ける耳。伝わる口。考える頭。

STEACは，日頃英語の音に接していない耳と口と頭を英語漬けにすることを狙った，夏／春休み毎日30分のオンデマンド特訓授業です。音声技術・言語技術，そしてAI技術を用いて，皆さんの「聞く」「話す」「考える」を鍛え，皆さんの能力を可視化し，スコア化し，評価し，その都度，フィードバックを返します。

夏休み：工学部３年生対象，春休み：工学部２年生対象（各１単位）

工学系及び情報理工学系研究科は2026年度から授業が英語化されます

学部生のうちに「聞ける耳，伝わる口，考える頭」を身につけましょう



Thank you, any questions?