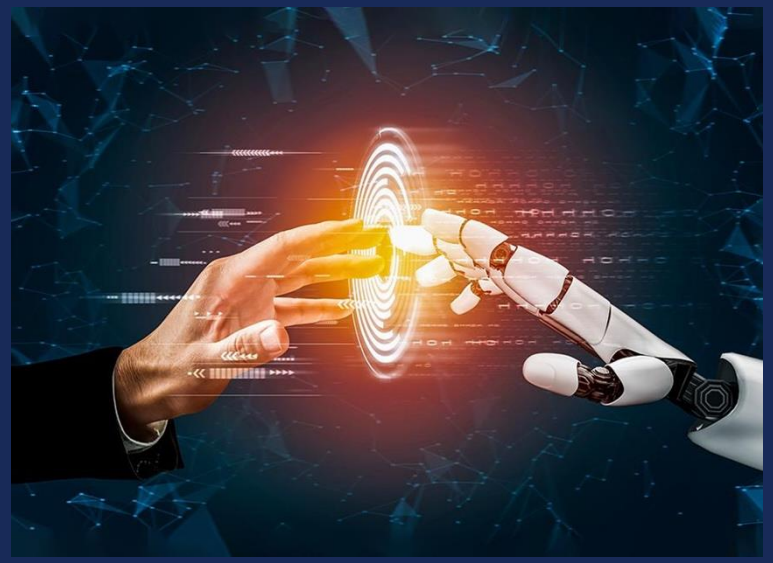
L2 Intelligibility in the Contexts of AI and Globalization



Okim Kang
Northern Arizona University

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

Tampere University

The Nordic Speech Research Forum
January 26, 2026

NAU NORTHERN ARIZONA UNIVERSITY

L2 Intelligibility in the Contexts of AI and Globalization

- L2 Speech and constructs
  - Overview of L2 speech technology and AI
- Critical AI in L2 speech
- Intelligibility in global contexts
- Future directions

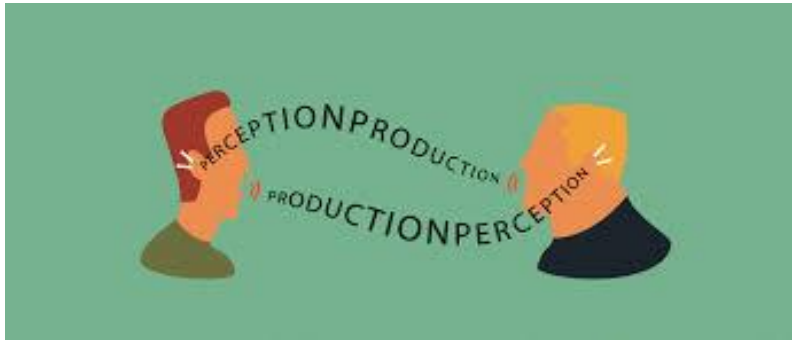
NAU, Flagstaff, Arizona

Altitude: 7,000 feet

A 90-minute Drive from Flagstaff to the Grand Canyon

L2 Speech & Accent

- **L2 speech:** How speakers **perceive**, process, understand, **pronounce** the sounds of an L2 language (Flege, 1995)



- The sound of speech is a critical issue in language assessment because people tend to **immediately judge native/nonnative speaker status on the basis of pronunciation (accent)** (Luoma, 2004).

- **Accent** is "phonological difference from a local norm".
- **Indeed, we all have accents!**
- **Accent is dynamic**: Accent modulates as speakers negotiate social identification or distance with listeners (Giles 1991, Communication Accommodation Theory)

New Perspectives in Global Contexts

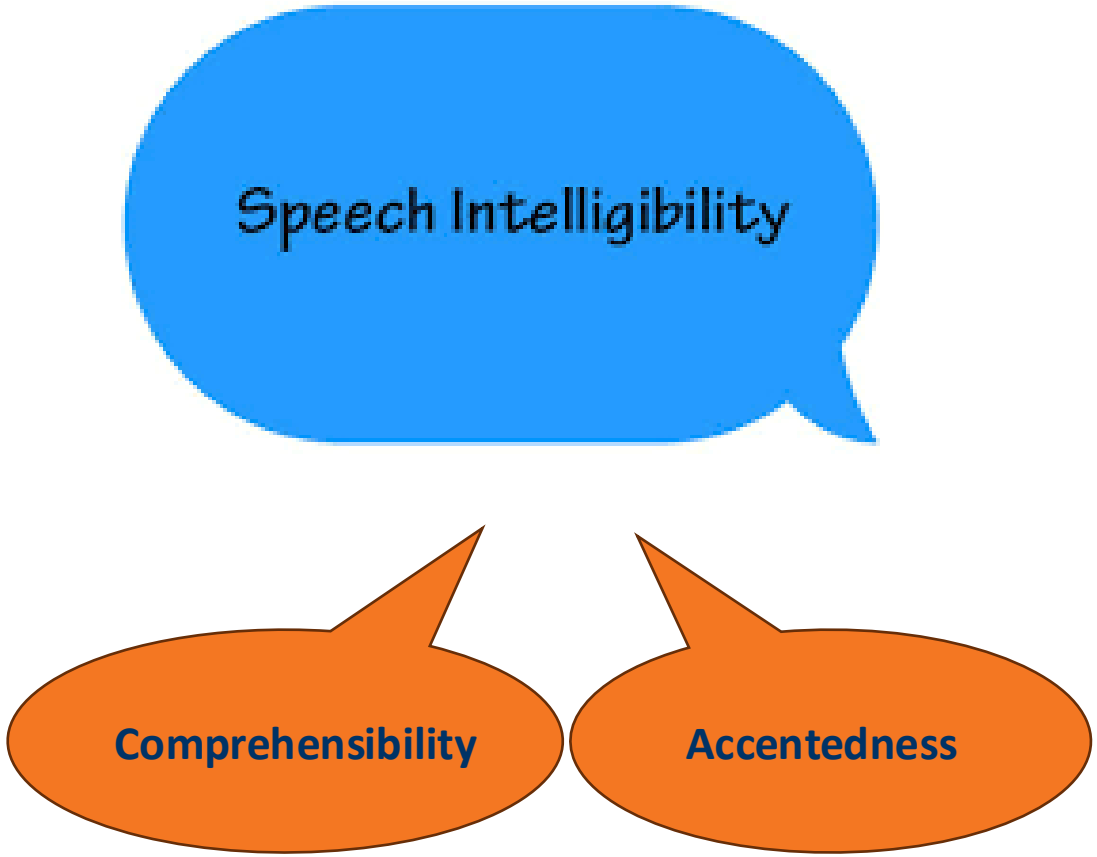
- English has spread all around the world. through globalization (Rose & Galloway, 2019)
- **Ownership of English** no longer belongs just to the inner-circle countries (Kirkpatrick, 2008)

- **Global intelligibility** is a key issue for successful communication in international contexts (Jenkins, 2006)
- **Global Englishes:** Broadly reflecting the use of English across borders and incorporating concepts of WE, ELF, EIL (Galloway & Rose, 2005)

Speech Constructs (Derwing & Munro, 2005)

- **Comprehensibility:** Perception of how easy it is to understand an utterance.
- **Accentedness:** Perception of how different a speaker's accent is from that of L1 community.
- **Intelligibility:** how much the listener actually understands of the intended message.

-Acceptability, appropriateness, familiarity, etc.



What is Intelligibility?

Speech science

- Defined as “the ensemble of properties that allow a native listener of the language to **correctly recognize the linguistic units (such as phonemes, syllables, and words)** in the order that they were produced by the speaker of the utterance” (Gooskens et al., 2010, p. 1022)
- Speech clarity

- Being **intelligible** is not synonymous with being **accent free** (Goodwin, 2014).
- **Mutual intelligibility:** Intelligibility does not reside solely in the speaker or the listener, but rather in the interaction between the two (Kang et al., 2014; Smith & Nelson, 1985).
- Intelligibility can **vary according to audience** (Brown, 1991)
- Intelligibility is **relational, co-constructive, and context-specific** (Kang et al., 2025; Kang et al., CUP, accepted) *(Contemporary Issues in L2 Speech Intelligibility)*

L2 Intelligibility (Kang et al., LT, 2025)

- Global Contexts
  - **Mutual understanding** among speakers of different English varieties
  - Requires both speaker's and listener's ability to comprehend (Gooskens, 2017).
  - What constitutes intelligibility v**aries from context to context** (Kang & Hirschi, 2025)
  - Most intelligibility research focusing on human judgement (Kang et al., 2025)

- AI Contexts
  - A new era in speech intelligibility
  - With the rise of interaction with AI chatbots (Labadze et al., 2023). learners or speakers are now prioritizing intelligible speech in **communication with AI** (Moussalli & Cardoso, 2020).
  - The evolution of ASR capabilities in AI technology has spurred **ASR-based L2 intelligibility** (Inceoglu et al., 2023).

Overview of L2 Speech Technology and AI

(Dentaa, Bae, & Kang, forthcoming)

- Automatic speech recognition (ASR): Recognizing isolated spoken digits of **one individual's speech** (Bell Laboratory, 1950s; Davis et al., 1952) => the recognition to vowels and phonemes
- Large-vocabulary continuous speech recognition systems (LVCSR) + Hidden Markov Models (HMMs) => recognized more **natural speech** (Witt & Young, 1997).
- In late 1980s and early 1990s, applying speech recognition technologies to **language learning**
  - Computer-assisted pronunciation training (**CAPT**) systems
- In early 2010s, technological advances in machine learning, including deep neural networks (DNNs), recognizing **other English accents**.
  - ASR systems became available on personal devices (e.g., **smartphones and computers**)
- In early 2020s, advances in end-to-end neural ASR and self-supervised speech recognition stable for large-scale use on **mobile platforms and in classrooms** (Prabhavalkar et al., 2024).
- Between 2020 and 2026, ASR systems have been integrated with **Gen AI** based on LLMs, enabling feedback in the form of explanations, prompts, and **conversational responses** (Hono et al., 2024).

L2 Speech and AI Tools

- In 2020s, mobile technologies and **intelligent personal assistants** (IPAs).
- In ASR-LLM models, speech recognition output combines with GenAI systems to provide explanatory and **dialogic responses in natural language** (Goh & Aryadoust, 2025).
  - Voice-enabled assistants (e.g., Google Assistant, Amazon Alexa) and speech-to-text systems (e.g., Google Voice Typing).
- **Conversational agents** to improve fluency, i.e., chatbots to rehearse low-stakes speaking opportunities
  - OpenAI's ChatGPT voice mode, ELSA AI Tutor, Duolingo Max, etc.
- Lower-cost and widely available tools for formative feedback
  - Google Voice Typing, Google Assistant, Microsoft Dictate, and Whisper-based transcription tools
  - **Word error rate => intelligibility**

Human vs. AI: Linguistic Variation

In comparison to human/human, when speakers communicate with an AI agent, they make…

- Comparable IC features (Karatay & Xu, 2025)
- Louder with AI than human (Siegert & Kruger, 2022)
- Timpe-Laughlin et al. (2022)
  - Fewer turns with AI (half)
  - Longer speaking time with AI
  - Fewer backchannelling with AI
- Longer pauses between turns or different **intonation patterns** (Kang et al., in process)

Intonation (Human vs. AI)

- Prosodically, speakers often signal agreement with **pitch concord** (or matching) and disagreement with breaking (Brazil, 1998; Pickering et al., 2012)

A: It wasn't my fault, was it?
B: No, of course, it wasn't. Definitely not.

Disagreement mitigation with concord (Pickering, 2018, p. 65)

Pickering (2018)

Human

Lily

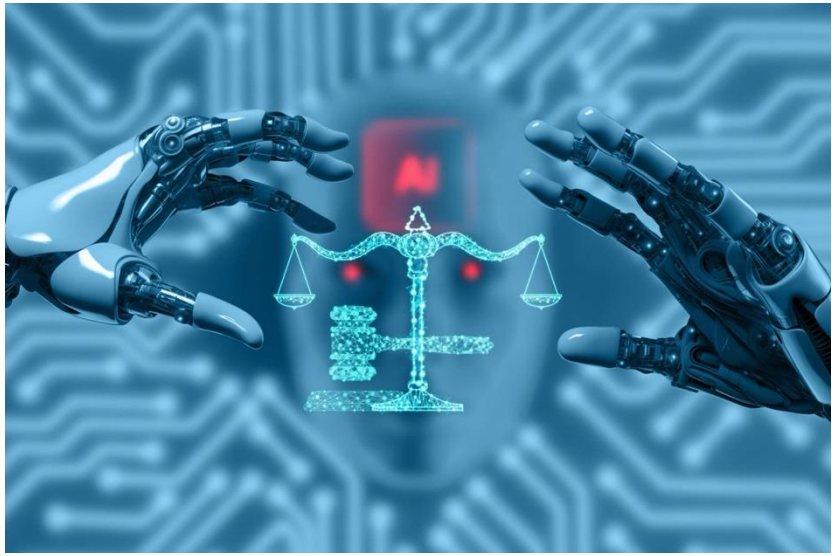
2025 version

Critical AI in L2 Speech

NAU NORTHERN ARIZONA UNIVERSITY

Kang, O., & Hirschi, K. (2025). AI-based bias and second language speech technology. *Annual Review of Applied Linguistics*.

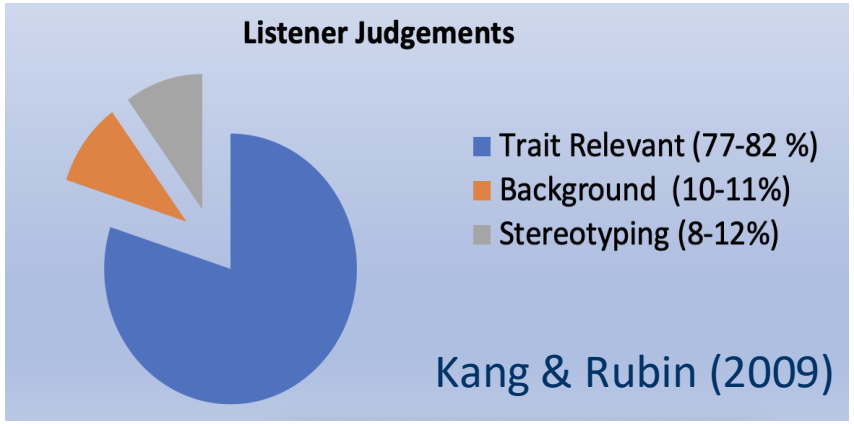
- **Does AI have any bias?**
  - Race/religion?
  - Gender?
  - L1s accent?



Human Listener Stereotyping and Background on Perceptions of L2 Speech

Many factors can affect our perception or judgement….

- Stereotyping (Kang & Rubin, 2009; Kang & Yaw, 2021; Levis & Moyer, 2014; Lindemann, 2002; Telo et al., 2024)
- Accent familiarity (Browne & Fulcher, 2017; Gass & Varonis, 1984; Miao & Kang, 2024; Ockey & French, 2014)
- Topic familiarity (Gass & Varonis, 1984)
- Shared-L1 effect (Kang et al., 2019; 2023; 2024; Munro et al., 2006; Shin et al., 2021)
- Training effect (Kang et al., 2020; Xi & Mollaun, 2011; Winke et al., 2013)
- L1 status (Fayer & Krasinksi 1987; Kang, 2012)



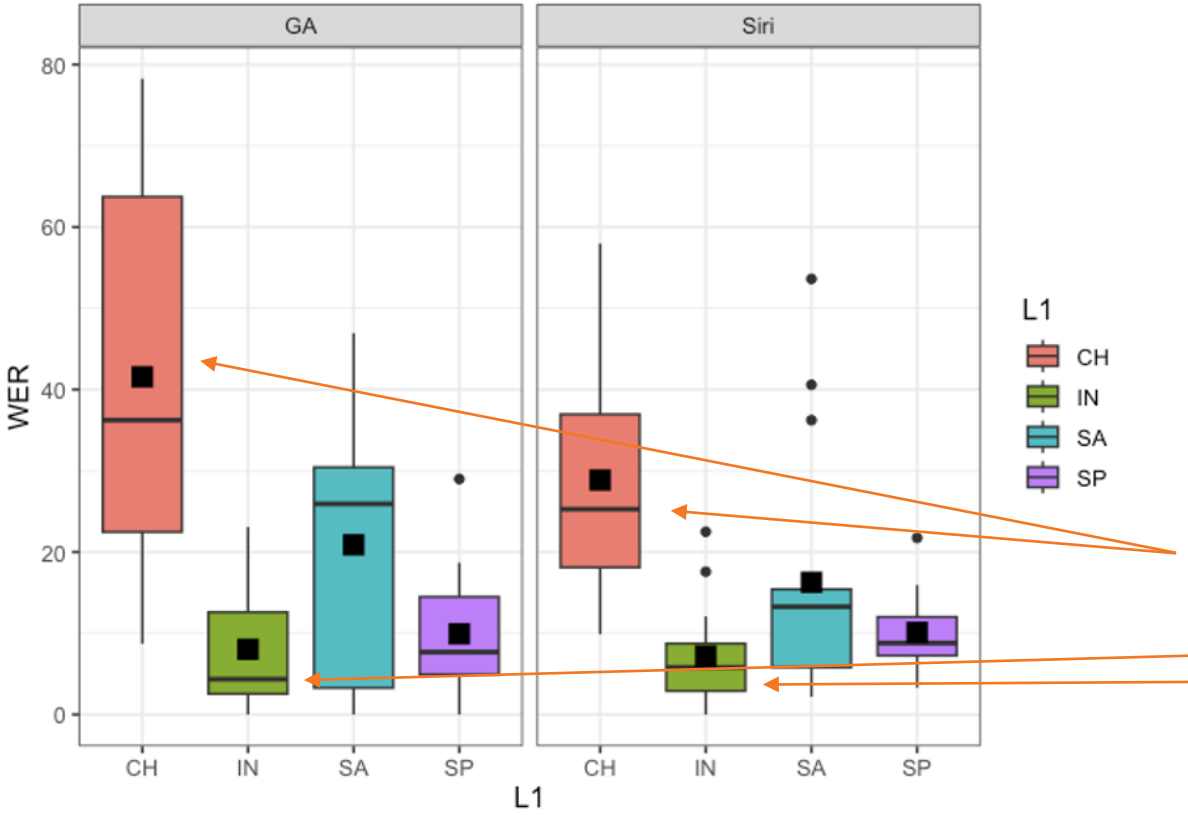
- Listener's attitudinal and background factors (Kang & Rubin, 2009)
  - **18-23 %** of variance

AI Understanding of Accent Varieties

- ASR transcription accuracy:
  - Persistent misrecognition of linguistic varieties (Choi & Choi, 2025)
  - Sustained error rate variability among 26 different English accents (Swain et al., 2025)
  - Lower accuracy in African American varieties than White American varieties (Koenecke et al., 2020)
  - Lower accuracy in women speakers than men (Tatman, 2017)

Testing AI's intelligibility: Do voice-activated AI assistants have L1 bias? (Bae & Kang, 2025, under review)

Figure 1. Distribution of WER by AI Type Across L1 Varieties in Dataset 1



Note. GA = Google Assistant, Siri = Apple's Siri
Note. CH = Chinese, IN = Indian, SA = South African, and SP = Mexican Spanish.

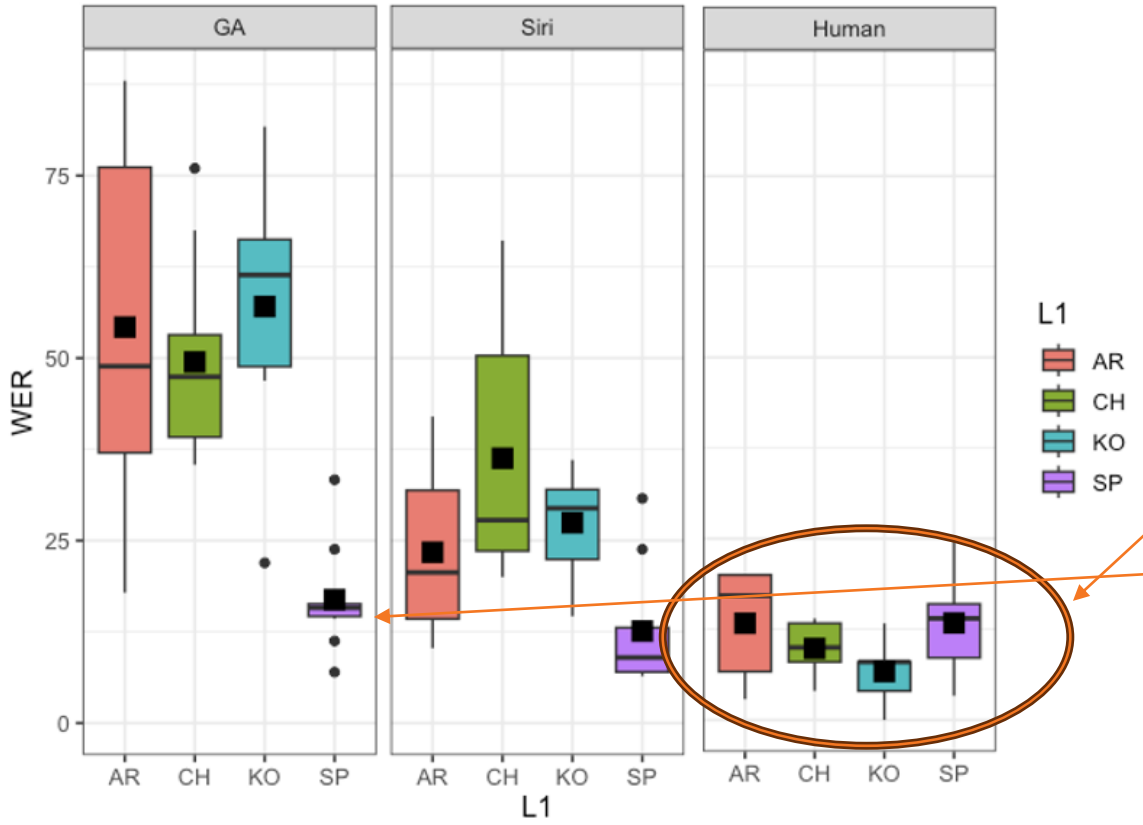
- **Dataset**: 60 Accent Archive speech
- Google Assistant vs. Apple's Siri
- **5 Expert raters** : High intelligibility confirmed through transcription (85% or above)

- Especially **Chinese accent** with a high word error rate
- **Indian and Spanish accents** with low error rates

AI Ratings vs. Human Ratings

(Bae & Kang, 2025, under review)

Figure 3. Distribution of WER by Rater Type Across L1 Varieties in Dataset 2

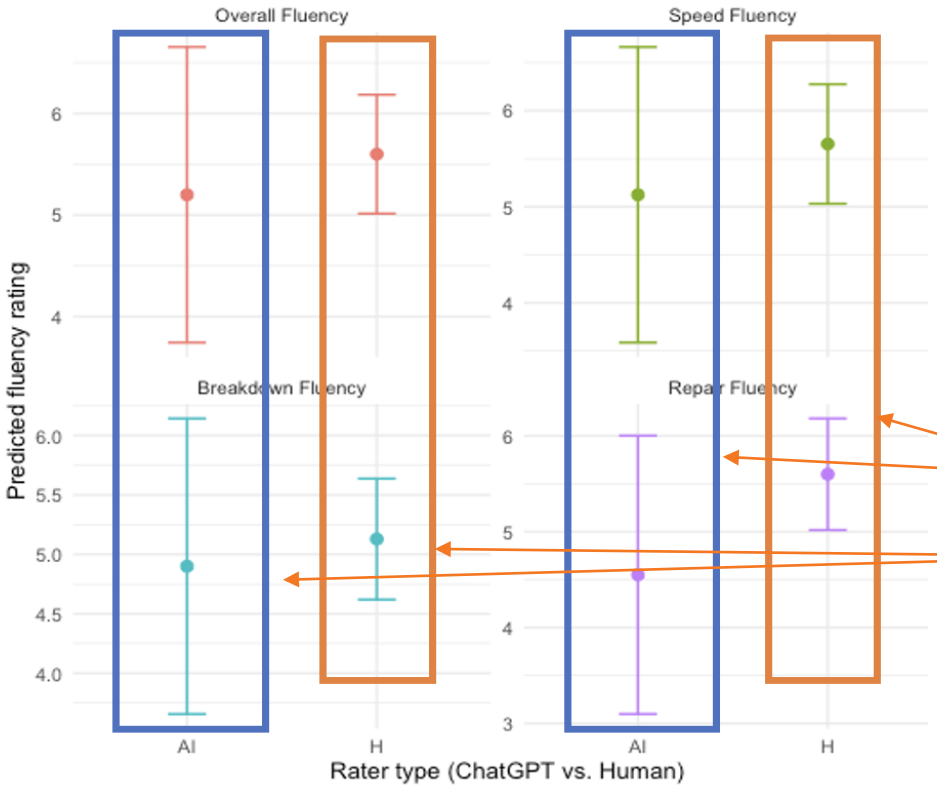


- Human raters are so consistent
- The **Spanish** accent has a low error rate.

AI as Potential Learning Tools for L2 Oral Fluency? (Bae & Kang, 2025)

ChatGPT vs. Expert Human Raters

Figure 1. Visual plot of the LMEM results



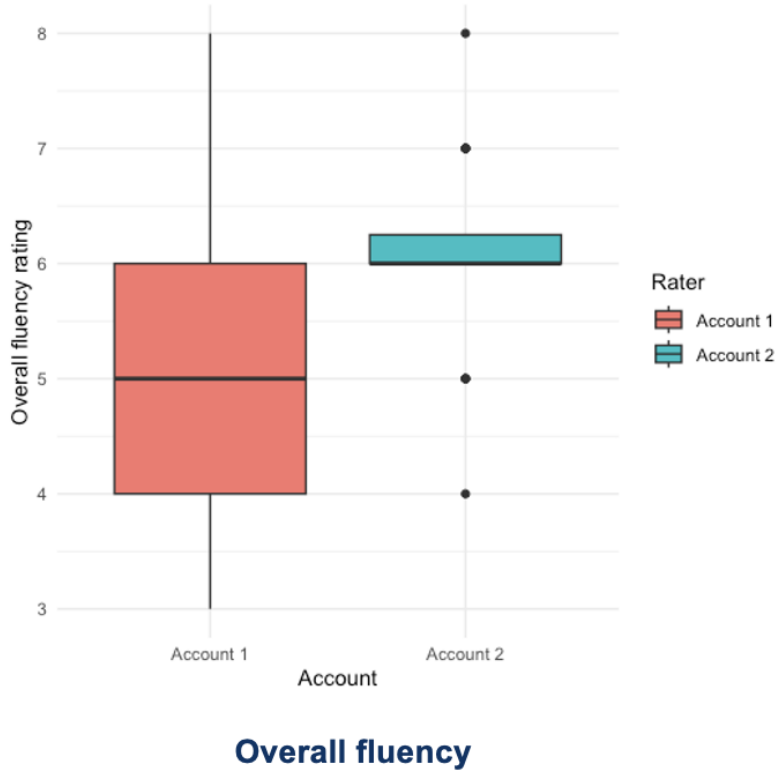
Note. AI = ChatGPT, H = Human experts

- 40 speech files from the Cambridge English Language Assessment (Kang & Yan, 2018)
- Human raters: 7 experts, ICC = .80 or higher)
- ChatGPT 4.0

- No statistical difference between the two rating types, but the clustering patters differed.
- ChatGPT: Broader variability

Inter-rater reliability: Consistency of ChatGPT fluency ratings across different accounts

Figure 2. Box plots of the results for each fluency feature rating across ChatGPT accounts

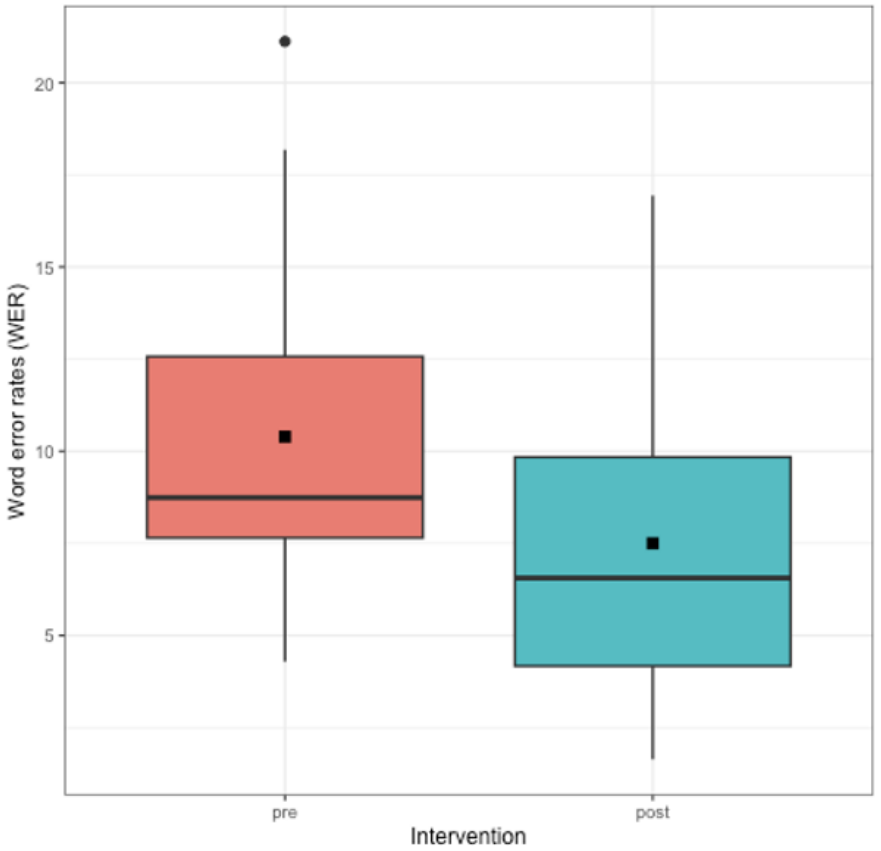


- This different account type explained 20% to 38.7% of the variance in ratings.
- A significant effect of account type (ChatGPT 1 vs. ChatGPT 2)

AI Training Through GE Varieties
(Bae & Kang, under review)

- **WE training** (Miao, Kang, & Meng, 2025) used for human listeners
- **The same training for ChatGPT 4**
  - Phonetic explanations and audio examples illustrating specific features of each accent variety for 30 speech samples
    - Chinese, Indian, Korean and Spanish
    - the phonological features in the accent varieties and differences from American English
  - Checked WERs

Intelligibility Scores



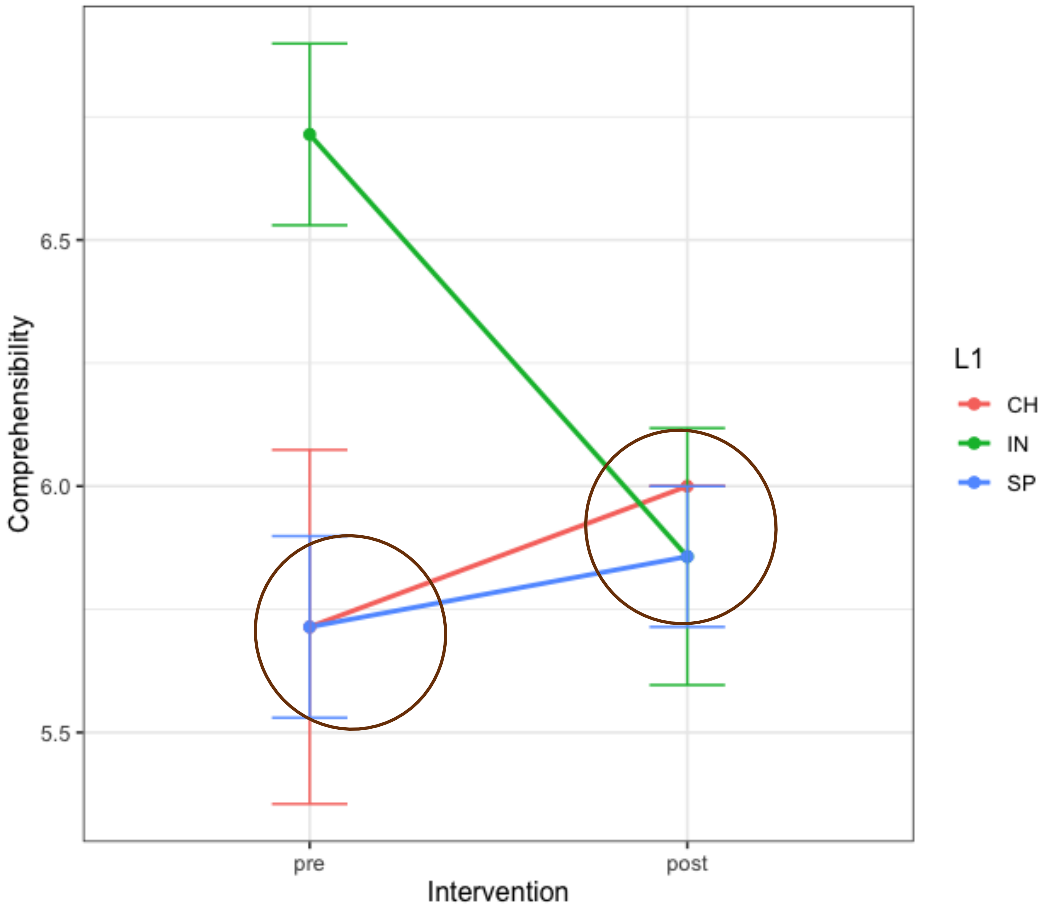
Pre=pre intervention; post=post-intervention

AI Training Through Cognitive Dissonance Theory-based Intervention

(Bae & Kang, in process)

Comprehensibility Rating

- 40 speech files from a human training study (Miao, Moran, and Kang, 2023)
  - WE phonological feature training
- Chinese accent perception improved (red line); Indian accent (green) dropped.
  - Differences merged into a much narrower variance



Human vs. AI Ratings of Personal Traits

(Hirschi, Kang, & Bae, 2025/ in-progress)

GPT Prompt

- How would you rate the speaker on a scale from 1 to 9 in the following descriptors?

The statement is…

- emotional—unemotional
- graceful—awkward
- plain—expressive
- effective — ineffective

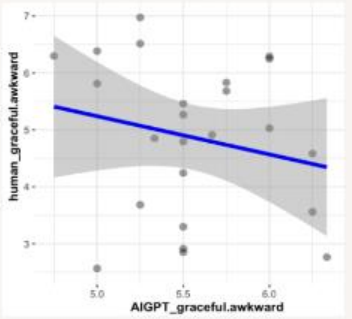
……..

- **The Study**
- 24 speech samples
- 161 listeners' ratings
- Preview-4o-Audio (Open AI API) model were given same instructions (April/2025)
- Compared them with human ratings

Very inconsistent ratings and no relationships

graceful/awkward

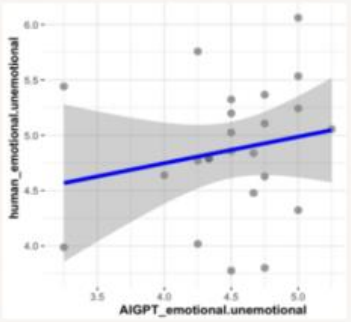
(r = -.21, p = .32)



ICC_{GPT} = .87

emotional/unemotional

(r = .20, p = .36)



ICC_{GPT} = .49

Intelligibility-Based Feedback: Objective Speech Intelligibility Measure (OSIM)

(Kang et al., ongoing)

- Create **a computerized system** for measurement of speech features and L2 speech intelligibility scores to provide teachers and learners with feedbac
- How doe
  OSIM co
  progress
  intelligib
  time?

There should be some kind of training for learners and teachers (instructional impact).

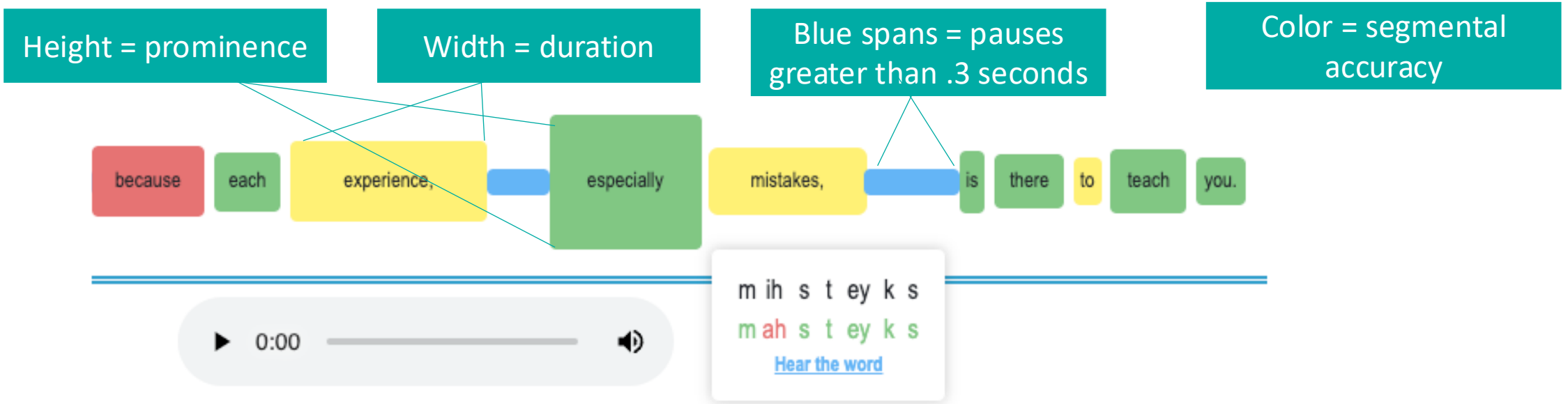
Speech analysis data:

Speech Event	Intelligibility	Speech Rate	Avg Pause Length	Rhythm Index	Word Stress Accuracy
Presentation Practice	97.67%	4.65 syl / sec	0.6 sec	0.6	79%

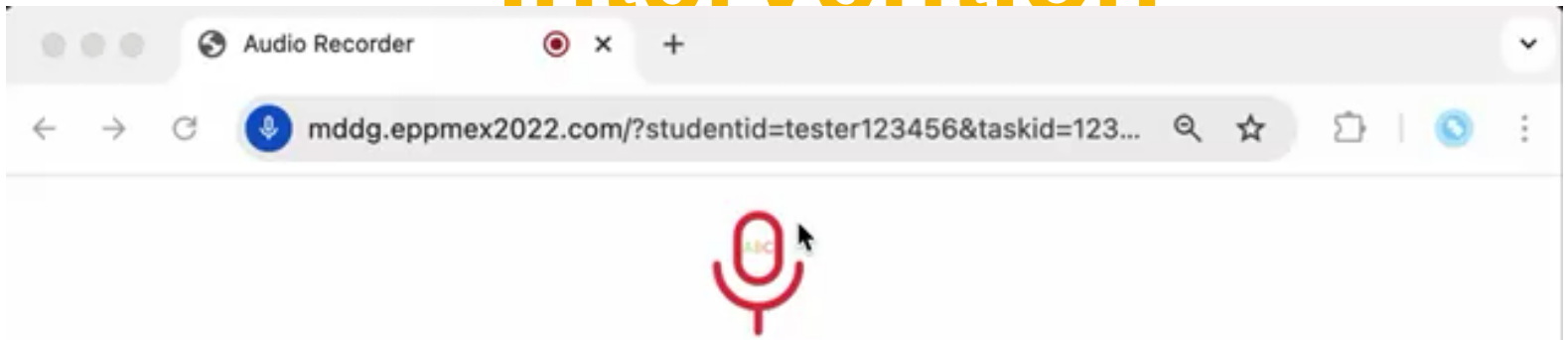
Your intelligibility is 97.67%	**Intelligibility**	ts, vowels, ake words difficult to understand. See the next page for a list of words to practice!
Your speech rate is 4.65 syllables per second.	**Speech Rate**	Your speech rate is near the target range (4 to 4.5 syllables per second). However, you can still practice speaking important words more slowly.

AI-based Mispronunciation Detection and ChatGPT Tutoring (Hirschi et al., LL, 2025)

- Fine-tuned wav2vec2.0 model (Yang et al., 2022)
  - ASR with Whisper, word-level segmentation with wav2vec2, fined-tuned wav2vec2 model which has a high accuracy rate, and prominence scores with Wavelet prosody toolkit, and finally ChatGPT feedback
- Phonemic-level error detection informed by L2 intelligibility



Intervention



Visual/Narrative Group: Significant improvements with ChatGPT tutoring in **pause placement**, and **prominence**

Critical AI

- Critical AI/digital literacy (Darvin, 2025)
  - The inequalities of user resources and the datasets of large language models (e.g., ChatGPT misunderstanding spoken Vietnamese)
  - The inequalities of high-resource and low-resource languages
- Access inequality: Some learners have more access to AI platforms than others
  - ChatGPT, Copilot, CharacterAI, ChatAI, Snapchat My AI, and Dall-E
  - **Paid vs. unpaid versions**
- Limited interpretability: Over-reliance without understanding each feature contribution (Zechner & Evanini, 2020)
  - AI-feedback situated clearly within construct
- Algorithms reproduce certain biases
  - Privileging some languages and forms of knowledge over others
  - **Transparency** (Buijsman, 2024): Users, educators, and learners being informed about system accuracy

Intelligibility in Global Contexts

NAU NORTHERN ARIZONA UNIVERSITY

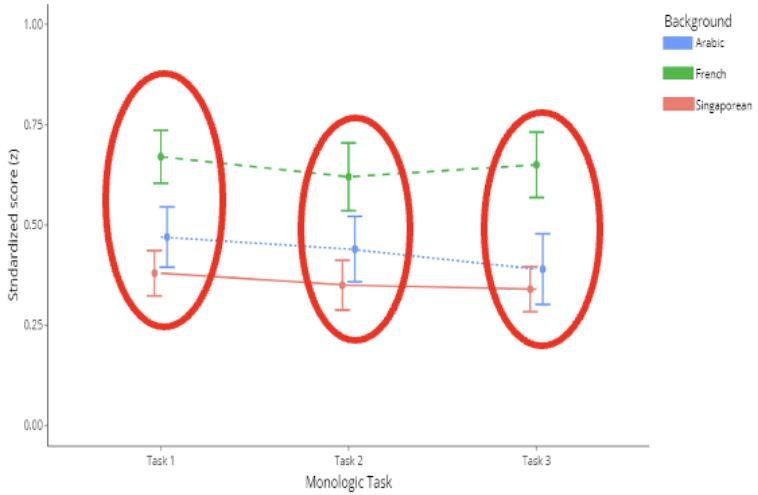
Learners' Perceptions Toward Accents vs. Their Listening Test Performance

Test-takers from different L1 backgrounds taking simulated high-stakes listening tests

- Kang et al. (2018): TOEFL
- Kang et al. (2022): the DET
- Kang & Kostromitina (2025): the British Council

No significant difference with highly intelligible speakers.

Kang & Kostromitina (2025): Aptis Listening test



L1 Shared L2 choice

Some emotional reactions: Korean test-takers hated an Indian accent.

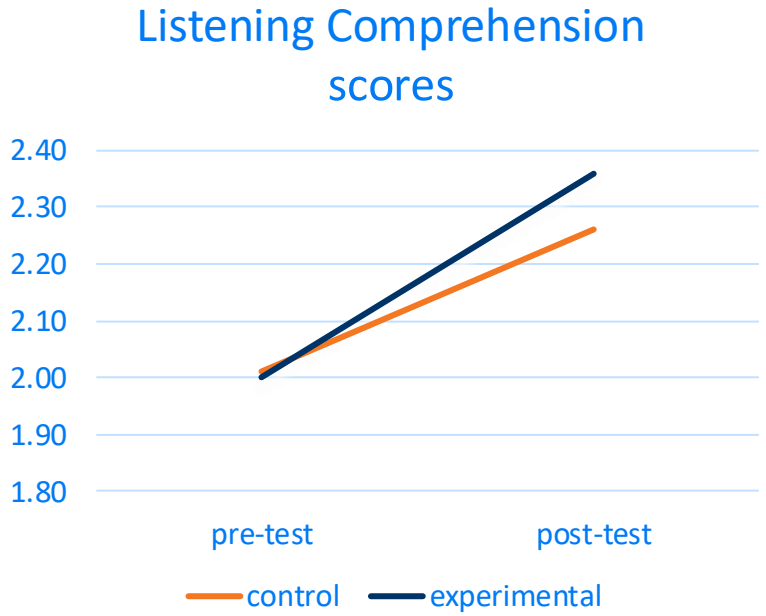
- Let learners choose an accent of their choice (Kang & Kostromitina, 2025): Ecological Validity

Different Accent Varieties

(Kang, Liu, Moran, & Miao, under review)

- Different varieties of accents for listening training (Kang & Liu, 2018)
  - 110 EFL students in Beijing
  - Listening skill training
  - **Control:** American accent
  - **Experimental**: Accent varieties

- High variability phonetic training (Barriuso & Hayes-Harb,2018)



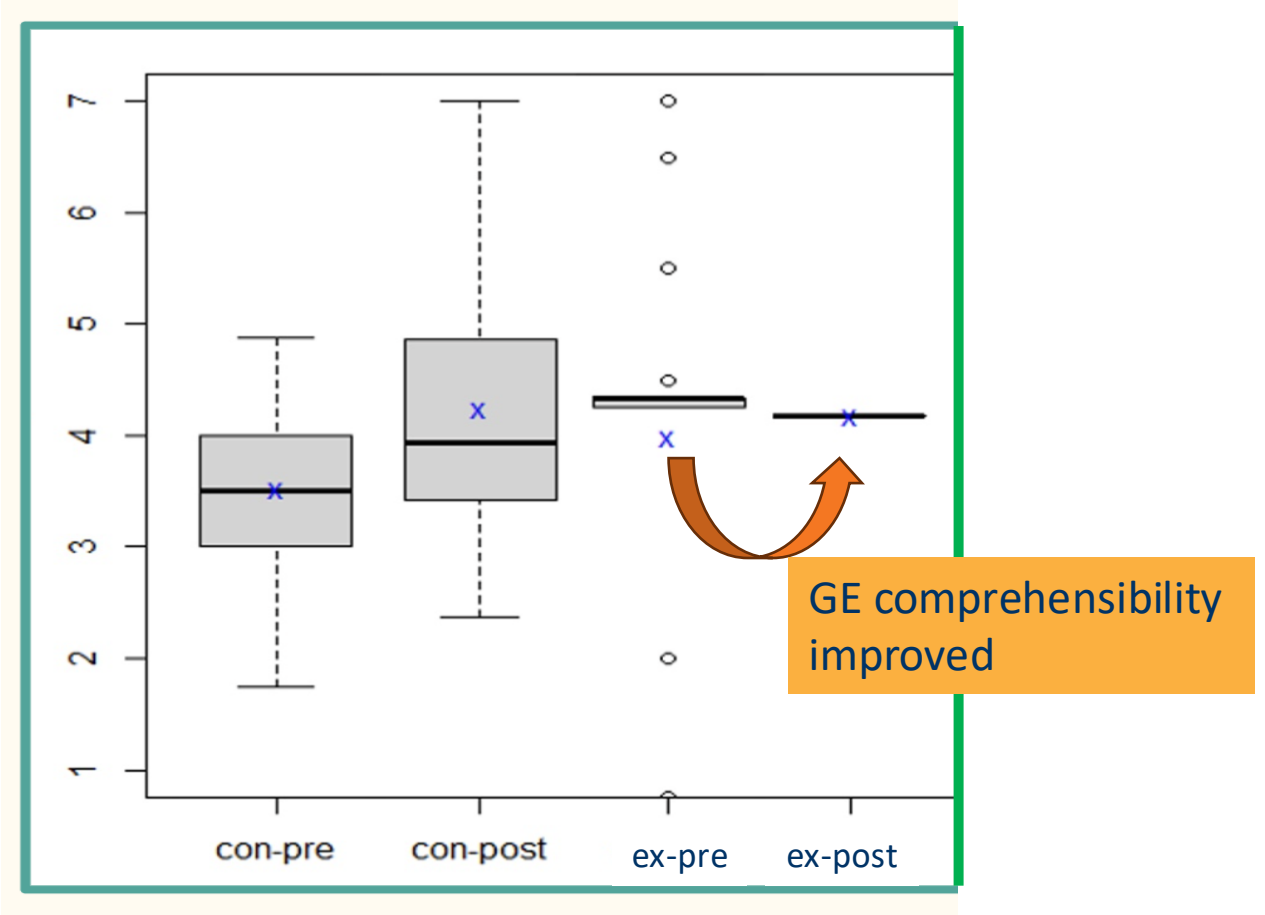
Only experimental group improved when learners listened to different accent varieties in their listening test.

Previous Approaches to Listener Training

Cross-cultural awareness	• Use of speakers' discussion of their experiences (Derwing et al., 2002)
Explicit instruction/ exposure	• Lessons on sound systems (Lindemann et al., 2016) or WE varieties (Kang et al, under review; Miao et al., 2025) • High variability phonetic training (Thomson, 2018)
Social psychology approaches	• Positive contact with L2 speakers (Kang & Moran, 2015; Kang et al., 2015; Yaw & Kang, 2021) or Dissonance theory (Miao et al., 2025)
Accent familiarity training	• Various accent-related measures and activities (Miao & Kang, 2023; Yaw, & Kang, under review)

The Role of Social Media Use in Listeners' Perception of Wes (Sullivan, Anolin, & Kang, in process)

- 24 participants for 2 weeks
- Follow and watch selected 10 videos from SNS influencers with GE accents
- Complete attention checks to ensure engagement
- Follow the platform's algorithms



GE Contexts: “Together We Communicate”

- Willingness to communicate
- Willingness to listen
- Listening skills
- Perceptions of social support
- Perceived common humanity, interpersonal skills
- Empathy

NAU Project: “Together We Communicate”
Kang (2024):

Future Directions

Future Research Issues

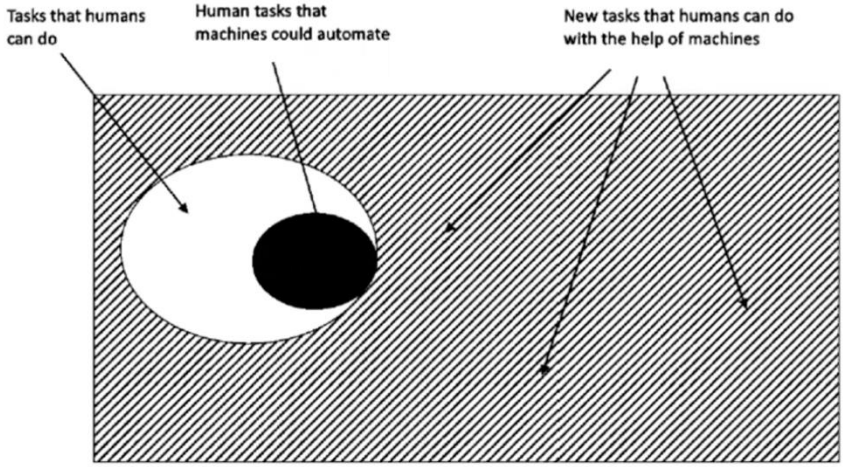
- AI cannot capture the full **complexity and expressiveness of human speech**, particularly in nuanced prosodic variations (e.g., **pragmatic and linguistic function**, etc.)
- L2 speech in global contexts
  - Pragmatically and culturally sensitive AI
  - L1 specific AI
  - Design for customizability: Let users select **target models**
  - New technologies (e.g., ASR, NLP, and AI) still largely rely on “native speaker” models

Future Research Issues

- Diversity, Equity, and Inclusion
  - Technology availability in all types of learners for all places.
  - **Critical AI: AI biased equity issues**
  - **=> AI trainer …human after all**.
- **Avoid one-size-fits-all model**
- Combine AI output with **human feedback**
- Collaboration among interdisciplinary researchers

GenAI Technologies for L2 Speaking: Critical AI

- Critical AI approaches:
  - Examining the extent to which the tools recognize their L1s
  - Comparing different platforms
  - Asking the GenAI platform for feedback
- There so many new tasks that humans can do with the help of AI.



Erik Brynjolfsson, The turing trap: The promise & peril of human-like artificial intelligence (Retrieved in January 2026)

Okim.Kang@nau.edu

Nordic Speech Research Forum