

# Variation in psycho-emotional impacts: underrepresented dialect speakers' experiences with Automatic Speech Recognition

Alicia Beckford Wassink

Department of Linguistics, University of Washington

<https://depts.washington.edu/sociolab/>

5 November, 2025  
University of Michigan  
NWAV53 Plenary Address



**NWAV53**



**SOCIOLOGUISTICS,  
CONFLICT, JUSTICE, PEACE**

# Bias in Voice Recognition

“They have Hoover Sparse sponsoring it.”  
They have whoever sponsoring it.

ASR is less accurate for certain speakers and  
speech communities



# Outline

## Part 1:

1. “What do I mean by ‘racial bias’?”
2. Acoustic sociophonetic study of ASR Errors that reflect bias in transcription of 4 sociolects
3. Wassink et. al. (2022) study of ASR Errors

## Part 2:

1. Study of psycho-emotional Impacts of ASR usage
2. Multi-city sample
3. Thematic Analysis
4. Results
5. Conclusions





# Collaborators

## University of Washington

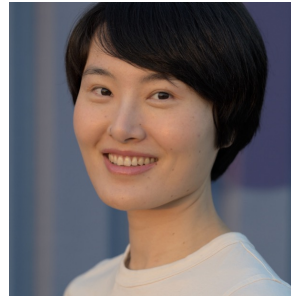
Anna Baptista	Siyu Liang
Isabelle Bartholomew	Helen Liu
Devyn Brandt	Katie Lindekugel
Jay L. Cunningham	Paige Morris
Reece Enders	Kat Ong
Megan Fajardo	Yadi Peng
Chase Fossen	Amaya Saunders
Cady Gansen	Amina Venton
Ty Gill-Saucier	Adya Veeraraghavan
Flo Humbert	Max Vu
Mwinkuma Kuunapor	Zoey Zhou

## University of Boğaziçi

Prof. Didar Akar

## Seattle University

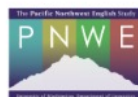
Dr. Didem Ikizoğlu



Affiliated with



COMOTION®





# Part I:

## Sociophonetic study of ASR errors

# Research Questions (Study 1):

Are there systematic differences in the performance of ASR on ethnicity-related dialect features? **Quantitative analysis.**

- Explore the differences in transcription accuracy (Phonetic Errors) of one commercial firm's ASR system for 4 ethnic dialects of Washington state.
- Which dialects are hardest for the system?

**Goal:** Produce computational tools leveraging sociolinguistic methods to combat algorithmic bias.

# Claims of Breakaway Success of ASR

- Breakthroughs of neural models: using Switchboard dataset,
  - system word error rate = 5.8%
  - human word error rate = 5.9%

Xiong et al. (2017)  
Dorn (2019)

- However, WER studies show that STT performance degrades sharply on regional and underrepresented communities

Harwell et al. (2018)  
Koenecke et al. (2020)  
Tatman and Karsten (2017)

- UX research: lack of user trust in ASR technology

Cunningham (2022)

- African American speakers feel uneducated
- Undue burden to code-switch to successfully interact with the system
  - “Siri, cut on the lights.” (natural utterance)   “Siri, turn on the lights.” (code-switched utterance)
- Georgia man unable to utilize On-Star during accident aftermath
- Systems are supposed to work for “normal users and use cases”

Wassink, Gansen & Bartholomew (2022)



# Background

## What do I mean by racial bias?

Defined for organizations as...

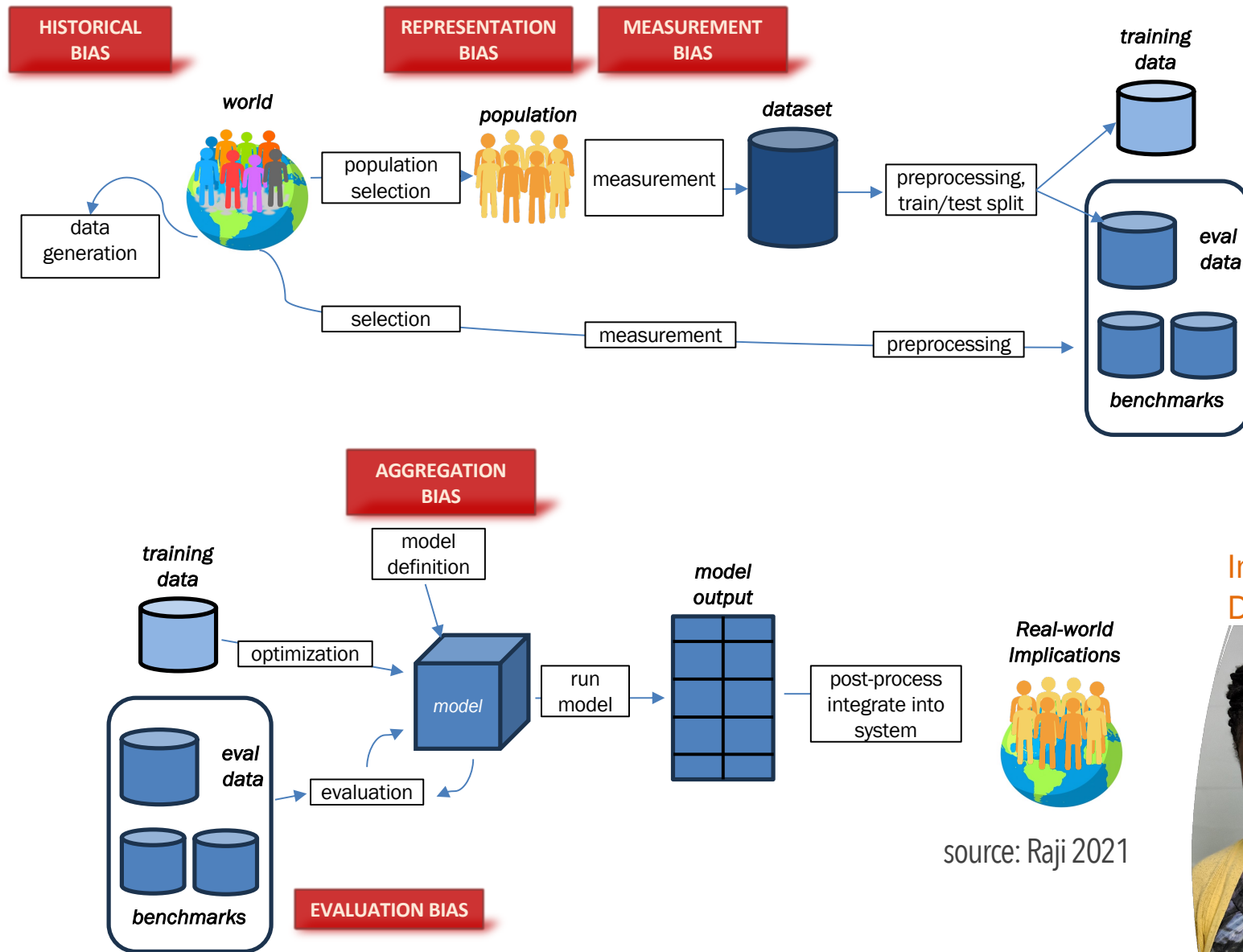
- 1) **Unequal access to the beneficial work of the organization,**
- 2) Racial disparities in the structure of the organization in roles and offices, 3) **Systematic patterns of inclusion and exclusion,** or hierarchical distinction, in how the work proceeds, 4) **Failure to examine disparities with intent to identify, address or reverse underlying causes.**

Maryfield (2018), Justice Research and Statistics Association  
Charity Hudley (2017)

The power dynamics that drive hegemonic, preferential treatment of people recapitulate bias. This is systemic racism at work.

Squizzero et al. (2021)

# Sources of Bias in ASR



Inioluwa  
Deborah Raji



source: Raji 2021

# Test Corpus: Pacific Northwest English

4 Ethnicity-related dialects

142 speakers

Each speaker: 3 speech tasks (casual conversation, reading passage, lexical task)

17 common phonetic variables

Leveraged regularities present in the structure of the traditional sociolinguistic interview.

## Tasks:

Casual speech

Lexical Tasks

Days of week, 1-10,

Semantic games

Reading Passage: “The Cat and the Mice”



Map credit: nationalatlas.gov ©2019: US Geographical Survey

# Targeted Sociolinguistic Variables



Code	Sociolinguistic Label	Example error	Target	IPA
Vowels				
(ɪ)	(ɪ)-tensing	peaking	picking	/ɪ/ → [i]
(ɔ)	caught/cot merger	com, cot	calm, caught	/ɔ/ → [a], [ɑ]
(æɡ)	pre-voiced velar (æ)-raising	beg	bag	/æɡ/ → [e:g]
Consonants				
(ɪŋ)	-ing (unstressed)	pick into	picking too	[ɪŋ] vs [ɪn] vs [ɪn]
(TH)	th-stopping	den	then	/ð/ → [d]
(l)	lenition	sheep	cheap	/tʃ/ → [ʃ]
(CC)	consonant cluster simplification	pace	paced	/st/ → [s]

# Phone Error Study Results

RQ1: Is there a difference in error rates between four ethnicity-related subsamples? **YES!**

- Overall nf, by ethnicity

Group	N=	nf
Caucasian American	19,142	1.6
African American	22,773	3.6
Yakama	22,695	6.3
Chicana/o	9174	6.6

One-Way ANOVA( $F(3, 788)=4.514$ ,  $p<0.001$ ). Tukey's HSD: Yakama~Caucasian-Am ( $p=0.04$ )  
Caucasian-Am~ChicanX ( $p=0.00$ )



# Part II:

## Psycho-emotional Impacts of ASR Failure



# Research Question (Study 2):

What are the human impacts of the failure of ASR to perform well for speakers from underrepresented dialects? **Thematic analysis.**

- Satisfied/unsatisfied expectations (presupposition analysis)
- Recollection of positive and/or negative memorable experiences
- Specific instances of dialect-related mistranscription

**Goal:** Collect data from the speech community to combat claims of Breakaway Success of ASR

# Methods

4 Research sites:	Atlanta, GA	Gulf Coast (MS, LA)	Miami Beach, FL	Tucson, AZ
Ethnicities:	Black or African American	French, Creole, Black and/or African American	Hispanic/Latine or Caribbean Heritage, Black or African American	Native American

## Elicitation conditions:

- Interview: recorded at location in target community
- Online: interview questions taken as survey (multiple US locations)

site ID	n=226	Ages	Gender Identity				
			F	M	NB	Prefer not to State	Trans-woman
Atlanta	97	18-60	24	71	2	-	
Gulf Coast	61	18-68	20	38	1	1	1
Tucson	25 (16 today)	18-33	19	6	-	-	
Miami Beach	43 (20 today)	18-60	31	11	-	1	

# Thematic Content Analysis

Def: a set of complex techniques used in qualitative research, allowing for **compression of large amounts of data into a focused set of themes**. Themes that emerge from the process may be used in **theory-building**. Categories are not supplied by a theory, but are established after preliminary analysis of the data, and over iterative passes (sometimes by multiple analysts) (Smith, 2000).

## Participants...

- make direct or indirect statements regarding the felt impacts of the use of ASR
- show us the frames of reference through which they view their interaction with ASR:
  - their expectations of ASR
  - their beliefs about those who use ASR and how it works for them
  - their feelings about how their language (repertoire) impacts their experience of ASR

## We, as analysts...

- utilize *in vivo* coding to extract information about experiences
- systematically code the data into themes using participants' own words
- suggest and incorporate new themes only when necessary
- replicate the study in a new location
- test and refine themes

modified Grounded Theory approach  
Glaser & Strauss (1967)

# Coding Procedures

## **Workflow:**

5 coders

2 analysts checked inter-rater reliability (Krippendorff's  $\alpha=.85$ )

Disputes discussed by team in weekly meetings over period of 1 year. If unable to reach consensus, settled by senior analyst.

**Coding:** Covered in next slide

**Summarization:** Custom R script extracted codes and their following sentence context from each question response:

1. Code and subtag extraction
2. Automatic aggregation into Themes
3. Manual checking of output
4. Additional graphical summarization used, per research question (e.g., Network graphing)

# Coding Procedures, cont.

**14 Codes:** (subset of larger coding dictionary used by the larger study)

- ATT - LANGUAGE ATTITUDE STATEMENTS
- DEV - DEVICE e.g., Siri, Alexa, Xbox Kinect
- FEL - FEELING WORDS
- GIC - GROUP-LEVEL IDENTITY CLAIMS
- IDE - IDEOLOGICAL STATEMENTS
- IND - INDIVIDUAL IDENTITY CLAIMS
- LAN - NEUTRAL STATEMENTS ABOUT LANGUAGE
- PSP - PRESUPPOSITION
- NAR - NARRATIVE
- BEH - BEHAVIOR (used alone or with the subtags below):
  - PSN - PERSONIFICATION
  - RPT - REPETITION
  - DISC - DISCONTINUATION
  - SHF - CODE-SHIFTING OR -SWITCHING

# Coding Example: PSP

## **Sentence 1. Mary won the game.**

A presupposition is the set of sentences that must be true for the target sentence (the proposition) to be true or false. If a conclusion can be drawn regardless of the proposition being true or false, they are presuppositions such as:

1. Mary exists.
2. There was a game.

## **Sentence 2. The <ASR device> recognizes my speech.**

Presupposes that:

1. The ASR device exists.
2. There is a speech event.

We may negate sentence (2) and the presupposition(s) still hold(s).

The <ASR device> does not recognize my speech.

1. The ASR device exists.
2. There is a speech event.

CD24-188 It is very [FEL: frustrating] because for [PSP: technologies [IGC: designed to make life better don't work for bi-lingual people of color.] Whereas [PSP: other non-poc around me have no trouble with the technology]]].



# The Survey

User Experience Survey (n=20 Total Questions)

- Demographic Questions
- UX questions
- Record a dialect sample

JotForm (allows embedding of self-recorded audio)

Delivered in person on-site or online (speaker's choice)

The screenshot shows a survey form titled "African American Users' Experiences in Voice Technologies". Below the title is a thank-you message: "Thank you for considering participation in our study! Take moment to share details and helps us get a better idea of it you're a good fit for the study." The form contains several fields: an "Email" field with a red asterisk and a placeholder "example@example.com"; a "Where in the U.S. do you reside?" dropdown menu with a red asterisk and a "Please select" option; a "What is your age as of today's date?" field with a red asterisk, a placeholder "e.g., 23", and a note "You must be at least 18 years of age to participate."; a question "Do you identify as Black American or African American and are based in the United States (U.S.)?" with radio button options for "Yes", "No", and "Other (further explain)"; and a question "If you would like, please share your gender identity." with radio button options for "Man", "Woman", "Gender non-conforming or non-binary", and "I prefer to not share".

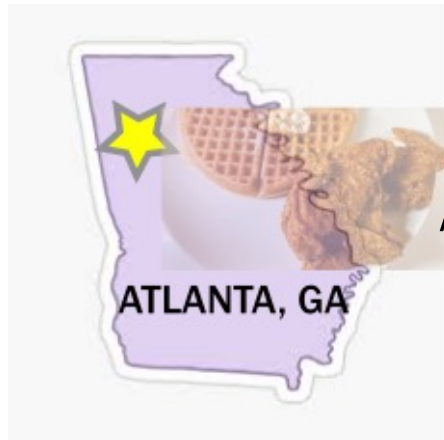
## 3 Key Questions (this presentation):

Q1) **How does it make you feel** when your voice assistant doesn't understand your speech/makes frequent mistakes?

Q2) Can you tell us about **your most memorable experience** with voice recognition technology? What was the technology and how did you use it? Did you accomplish what you wanted?

Q3) Can you **share a time** where you had a challenge with your **dialect or accent** being understood by voice recognition systems?

# How do you self-identify?



Black  
African American  
African (Kenyan, Eritrean  
Nigerian, Ghanaian)  
African-indigenous  
African-Hispanic  
Jamaican  
Cambodian-Black  
Hebrew-Black

Creole  
French  
White American  
African American  
Black  
Multi-ethnic



Hispanic/Latine  
Asian  
African American  
Black  
Caribbean

Navajo  
Dine/Navajo  
Chicasaw, white  
Tohono O'odham  
akimel oodam and Tohono oodam  
Yaqui and Mexican  
Other



# Feelings vs. Emotions

**Emotions** are the immediate, unconscious, biochemical responses to an external stimulus.

- physical reactions measurable in real-time
- e.g. Loud noise triggers “fear” response in the amygdala.

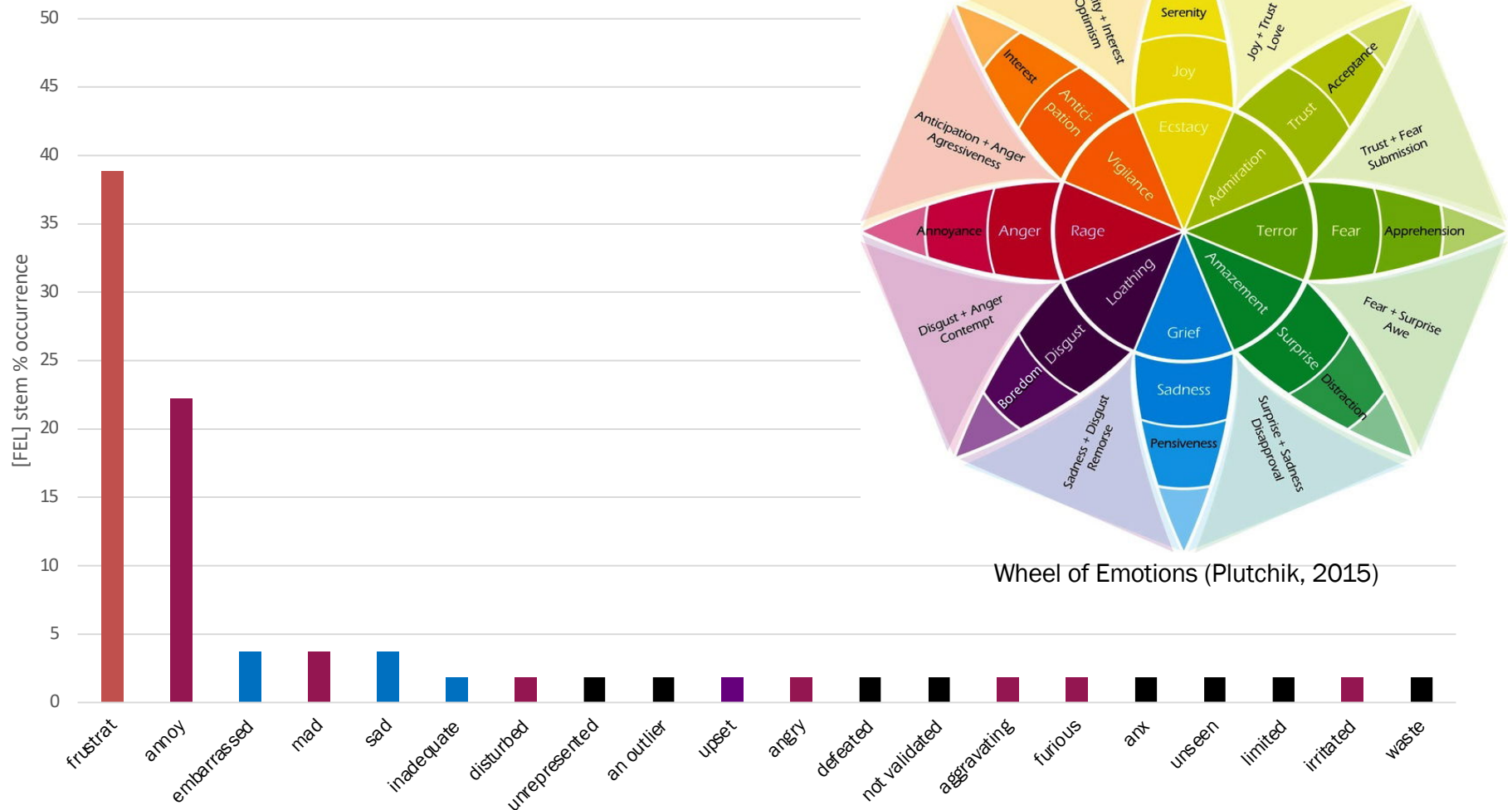
**Feelings** are the conscious, subjective interpretation of those emotions.

- mental awareness and label we give to a reaction,
- e.g., feeling "scared" when the loud noise triggers a physiological response.



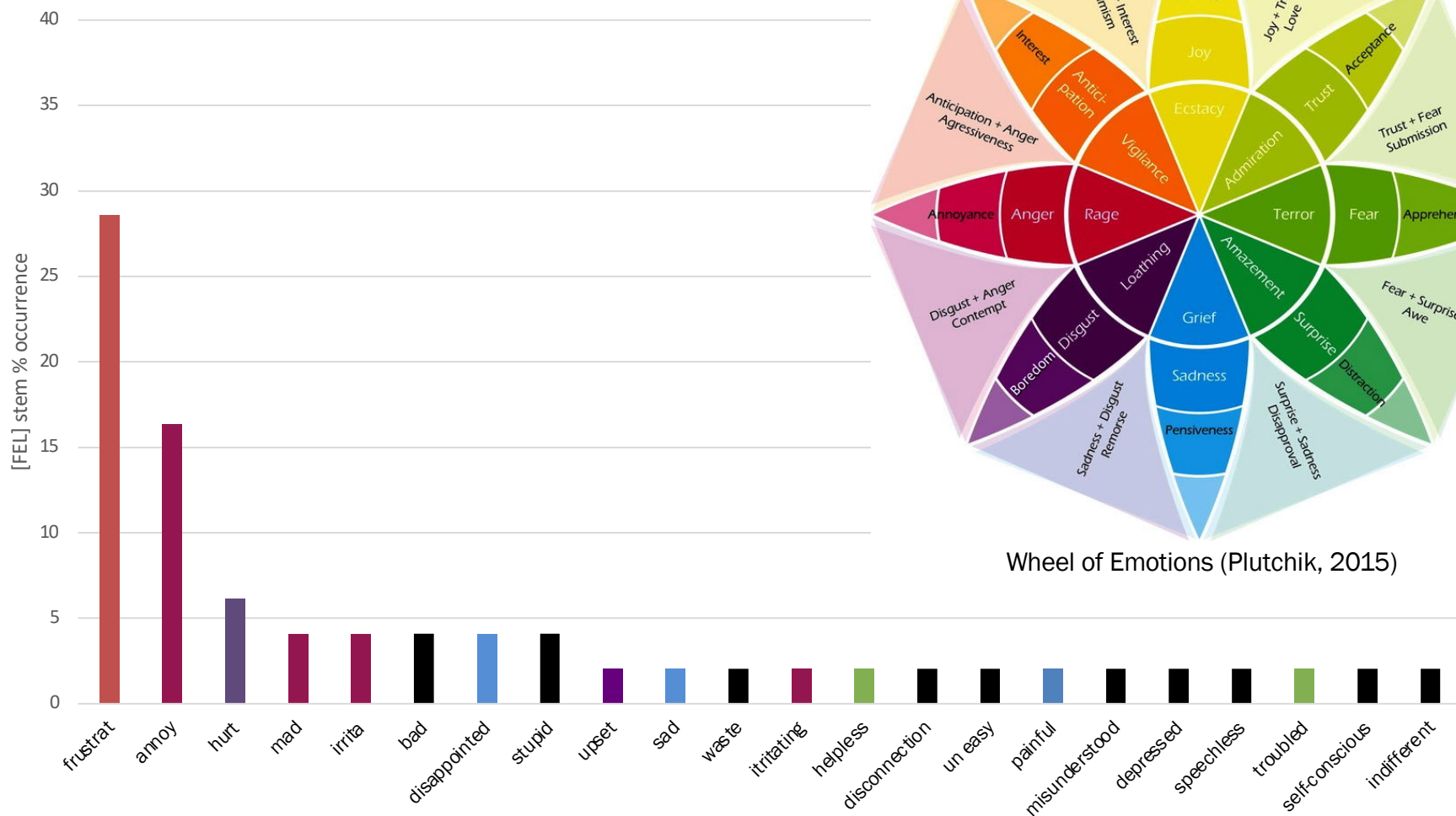
# Q1) How does it make you feel when your voice assistant doesn't understand your speech/makes frequent mistakes?

site: Atlanta (n=54 stems)



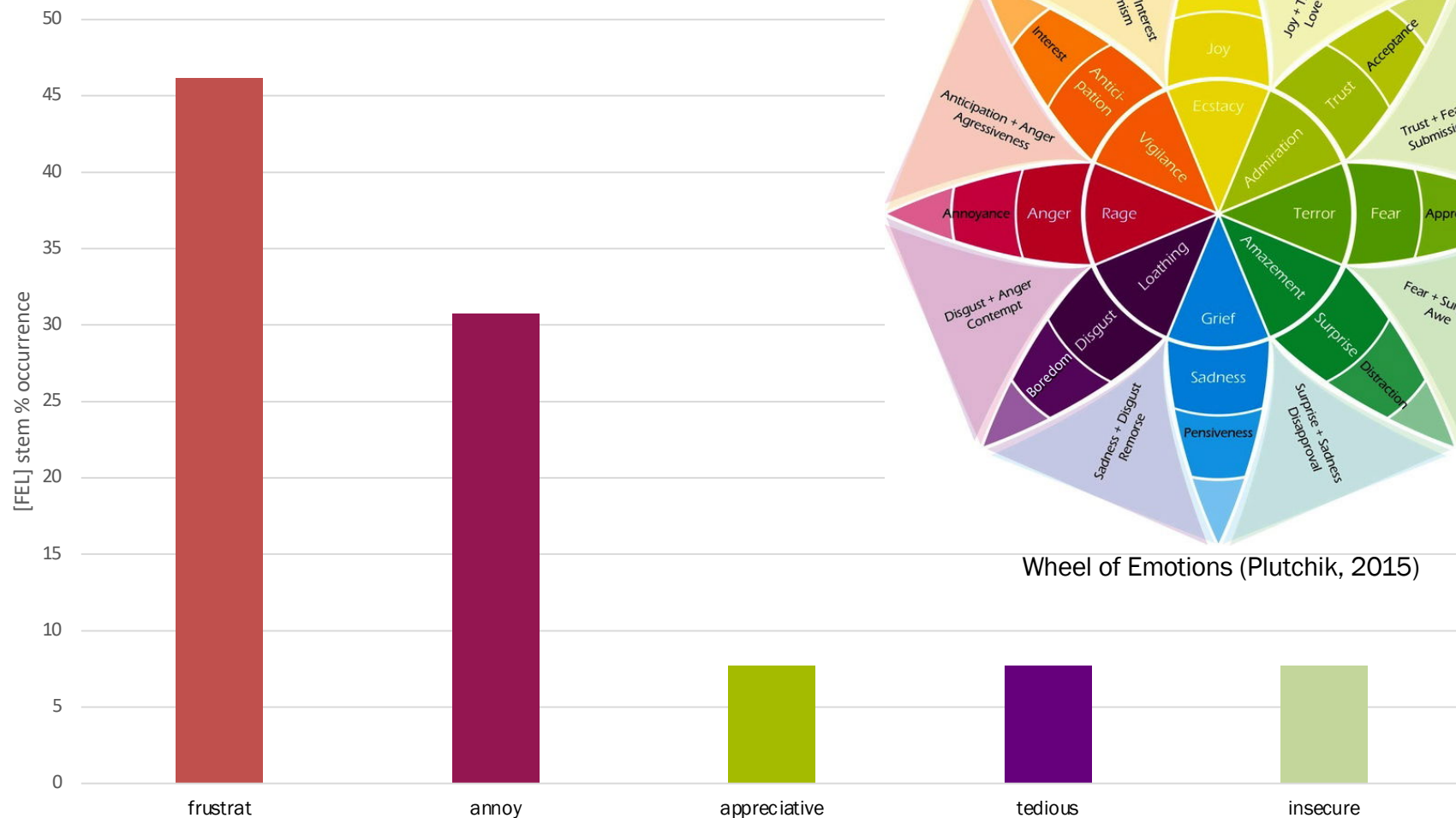
# Q1) How does it make you feel when your voice assistant doesn't understand your speech/makes frequent mistakes?

site: Gulf Coast (n=49 stems)



# Q1) How does it make you feel when your voice assistant doesn't understand your speech/makes frequent mistakes?

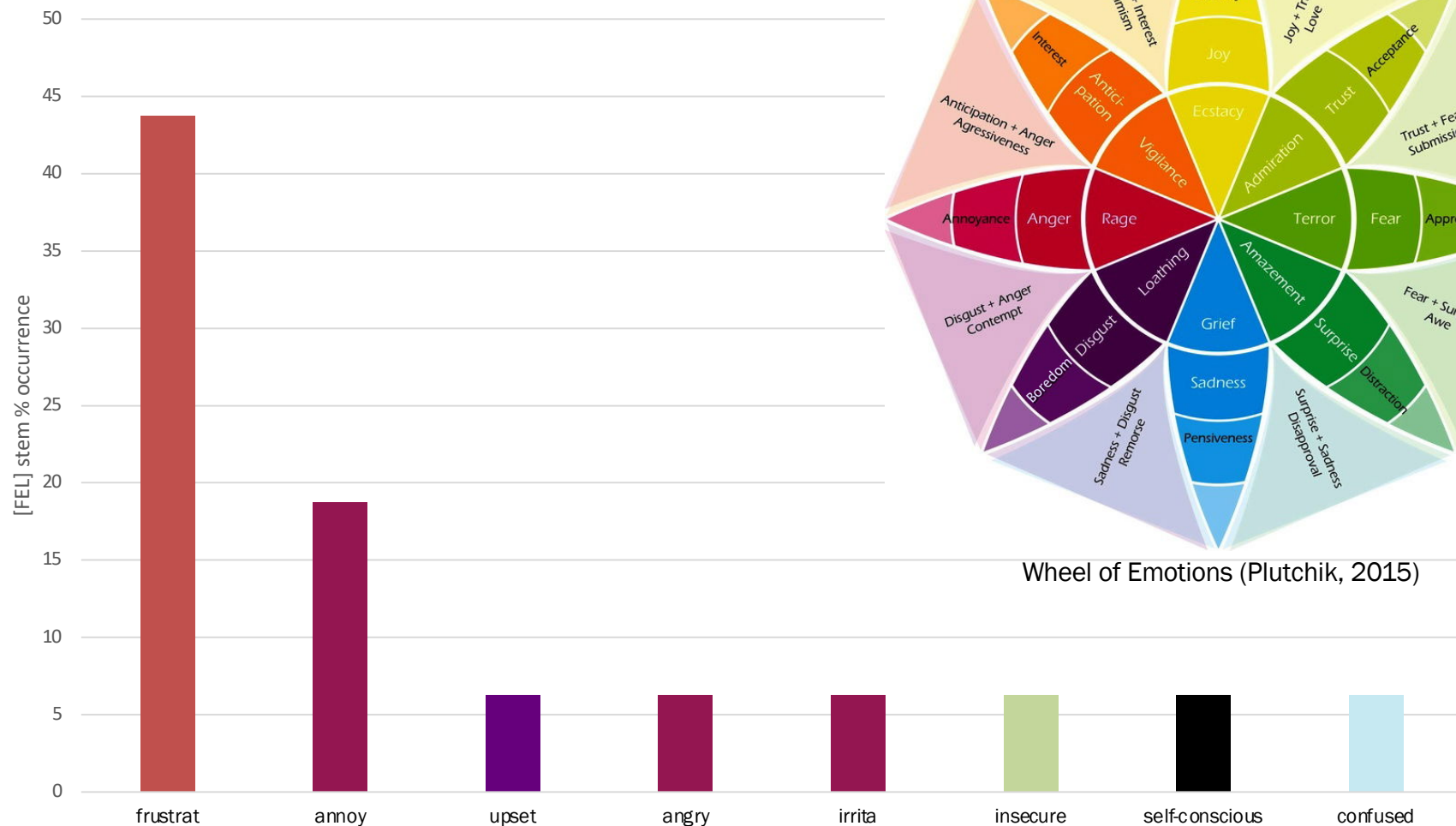
site: Miami Beach (n=13 stems)

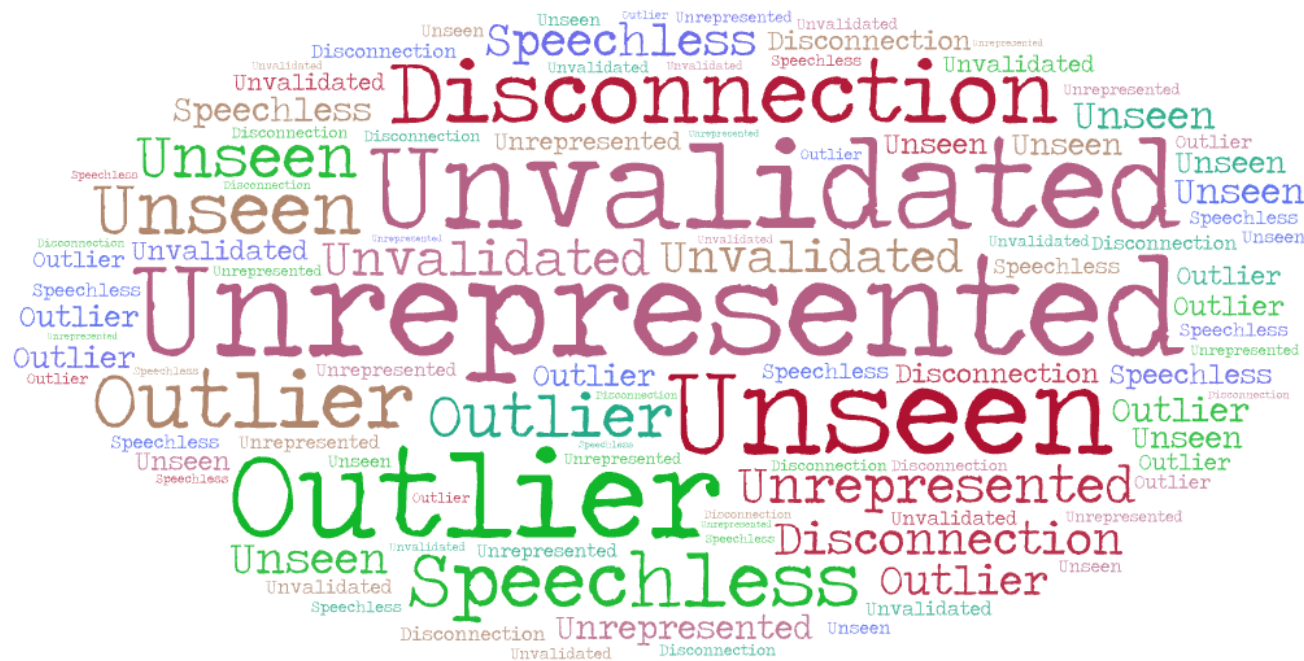




# Q1) How does it make you feel when your voice assistant doesn't understand your speech/makes frequent mistakes?

site: Tucson (n=16 stems)





CD24-606 (Navajo): [**FEL**: I can say that because [**LAN**: Navajo] is my first language I somehow feel an auto I get this automatic feeling of inferiority if that makes sense. Because you're I'm my in my mind I'm thinking this is not something that was made for me automatically. And **I'm stepping into a world of what other people are doing and not me because I'm native.**



# Invisibility: an emotion?

Invisibility is not an emotion, but a **feeling** caused by a complex of emotions, including:

Pain	Shame
Loneliness	Anxiety

Generalized feelings of invisibility are common: everyone at some point feels insignificant, unseen, unheard.

## Clinical statement (generalized feeling):

“Results from cognitive distortions rather than objective truths.”

Automatic negative thoughts (ANTs).

Amen (2025)

## Often may be managed by:

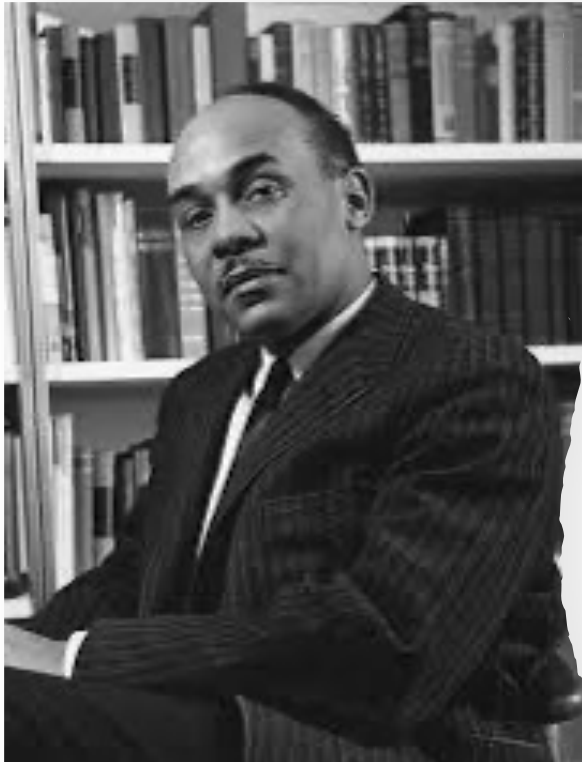
1. Questioning negative thoughts,
2. Learning to find validation from healthy sources,
3. Understanding brain connection,
4. Self-care.

# Invisibility Syndrome

## \*Intersectional invisibility

def: “Psychological condition produced when a person perceives that his or her talents and identity are not seen because of the dominance of preconceived attitudes and stereotypes.” (Harrell, 2000; Franklin, Boyd & Kelly 2008)

- experienced as a microaggression (Pierce 1995)
- can limit the effective utilization of personal resources, achievement or goals
- regularity of one’s experiences being questioned, requests for “proof”



Ralph W. Ellison, The Invisible Man (winner of 1953 US National Book Award)

“I am invisible, understand, simply because people refuse to see me.”

“And I love light. Perhaps you'll think it strange that an invisible man should need light, desire light, love light. But maybe it is exactly because I *am* invisible. Light confirms my reality, gives birth to my form.”

## Q2) Memorable experiences (negative)

Neutrality of prompt was intended to allow for positive, neutral or negative recollections.

(CD24-42) [NAR: ... [my nieces] asked me if they had permission to ask Google for "Moana music..." [PSN: The device... understood their cute baby voices] ... Later ... I went over and asked it to play "songs from the Frozen soundtrack". After a few seconds, we heard "I'm sorry I didn't get that," and then it proceeded to play Moana music.  
[BEH: [RPT: This rotation went on for a few minutes] of me getting [FEL: frustrated] ... **Something that started out being so simple, and meant to bring enjoyment, turned into an adult getting headaches, and two screaming children.**]]



(CD24-9) I don't think I have a most memorable experience. But **I always [FEL: dread] calling insurance companies for my job** because [PSN: the speech recognition system never understands me], and I end up getting transferred to a live representative. [BEH: I wish I could opt out of the recognition system because I end up in a negative feedback loop.]

(CD24-521Online) [NAR: It was an interview and [DEV: the AI] was the interviewee [PSP: **I was disqualified because it couldn't recognize what I was saying.**] I never accomplished and I had meet qualifications for the job]

## Q2) (Positive) Memorable experiences

(CD24-1):

[**NAR**: At the time I was hospitalized and had some school assignments due...I use my headset and was able to dictate two papers and have them turned in on time. Through technology **I was able to stay on task** with my educational requirements. **I was very successful** in accomplishing what I wanted to do through it.

(CD24-160): **It's a great experience.**



(CD24-18):

Transcription from speech to text is the most unforgettable experience. [**FEL**: **I like to use it**]. [**PSP**: It can help me finish it every time.]

(CD24-17):

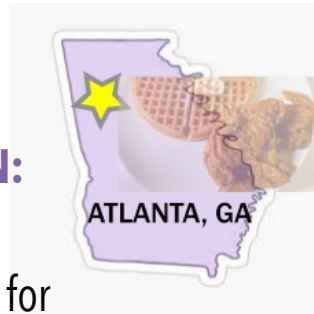
[**NAR**: I set up [**DEV**: Google Assistant] on her device, and [**PSP**: she was able to dictate the message using just her voice]. The technology **accurately recognized her speech**, and the message was sent successfully]



# Q3) Recall a time your dialect was misunderstood

(CD24-198) **I feel inadequate speaking my own first language.**

(CD24-14) **[PSP: Every time I use [LAN: my natural dialect] I have problems].**  
**[BEH: I have learned to adjust my dialect for [DEV: voice recognition systems]]. [ATT: I wish I could just speak normally].**



(CD24-9) **[PSP: [ATT: If I do not speak super proper] [DEV: Siri] doesn't understand me]. [PER: Most of [GIC: my names with black cultural roots] aren't understood when I ask [DEV: Siri] to call them] so [DEV: siri] will call the wrong person.]**

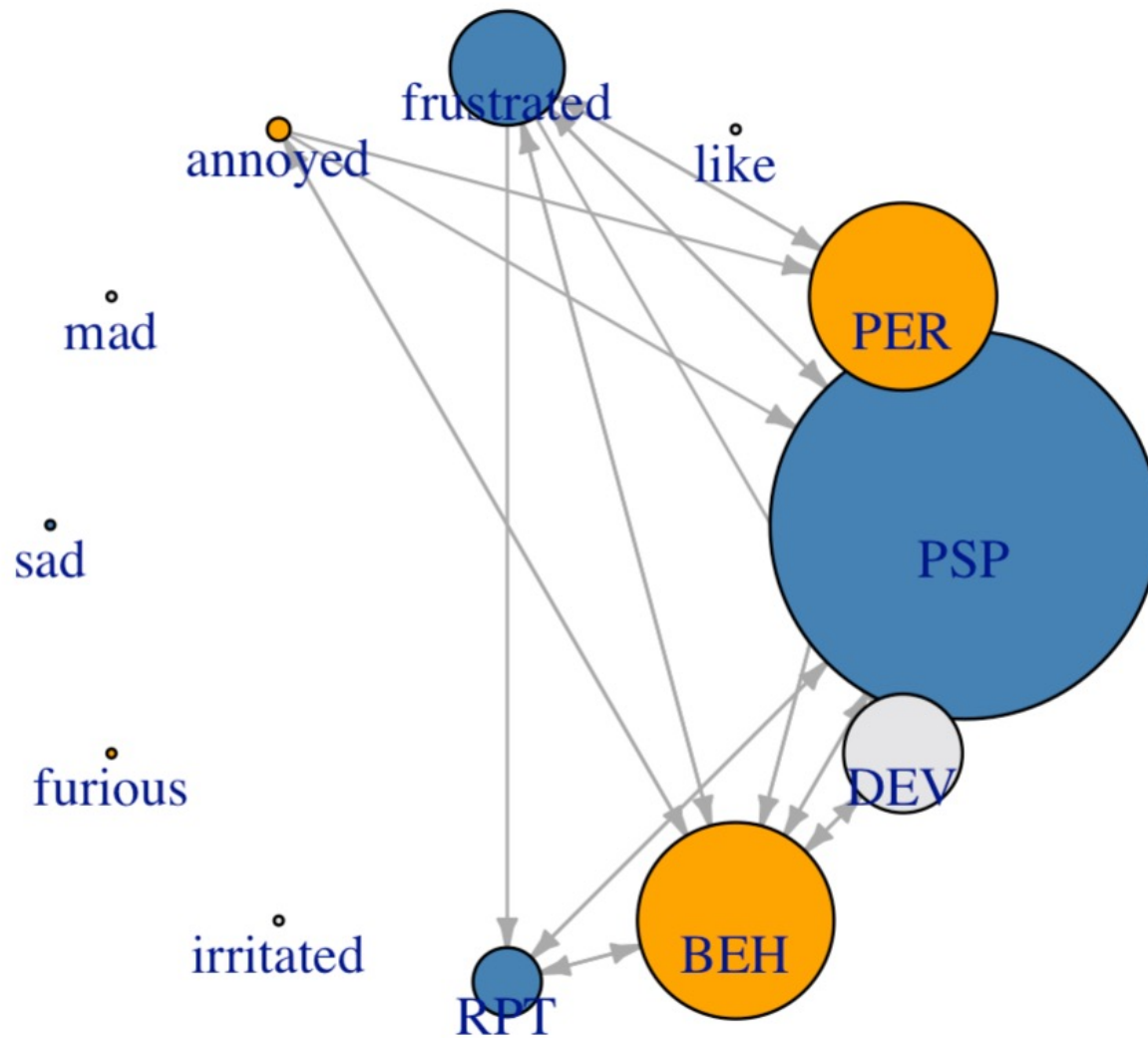


(CD24-578) my family, they have very strong accents even though they are fluent. they'll ask **[DEV: Alexa]** to do something, and **[PSN: she's like 'sorry I didn't quite catch that' or like 'please repeat that.'** **Like, girl's always lost. So yeah, she's kind of unreliable when it comes to like -- unless it's me, obviously.** But she's kind of unreliable when it comes to, like voice commands 'cause she's always lost. My girl does not know what's going on, like ever."



## Geodesic distances: Feelings, Personification and Presupposition

CODES: [PSP] Presupposition [PER] Personification [BEH] Behavioral CHANGE [RPT] Repetition [DEV] Device



# Personification

Miami, Guyanese F: Yeah yeah there are sometimes that I would ask [DEV: Alexa] something and she could pick it up before the first thing is her reply I'm sorry I don't know that **So I think we we need a Caribbean Alexa.** like [PSN: **an Alexa you know who knows Alexa talks because you know in the Caribbean we have our own [LAN: Our own Creole] And so an Alexa Caribbean is going to know the [LAN: Caribbean Creole]** because if you ask Alexa something that is Caribbean based she doesn't have that culture nor ethnicity that she can relate to So she's gonna say that obviously "I'm sorry I don't know that"



# The Impact of ASR Failure

- Societal bias is recapitulated in tech
- Non-white people feel voice assistants are not “made for them”
- Psychological harm: anxiety, anger, feeling “othered,” internalized sense of failure

Mengesha et al. (2021)



- Mild to severe consequences are possible:  
Failure of safety systems (recall OnStar example)

# Conclusions (Part 2)

1. Thematic analysis yielded rich data regarding repeated experiences and felt impacts.
  - Invisibility unexpected, but common
2. Extended narratives emerged despite survey format.
3. Analysis of presuppositions exposes respondent expectations regarding device/ASR performance.
4. Respondents show awareness of the burden they face to modify their speech away from what is “authentic.”
5. The “breakaway success” of ASR is not enjoyed by all. Certain individuals endure mild-to-severe impacts, not the gains promised to all “normal” users.

**frustration  
annoyance  
anger  
depression  
defeat  
invalidation  
a sense of feeling left out  
“speechless(ness)”  
an undue burden to code-switch  
required to employ  
increased effort**

# Thank you!

wassink@uw.edu



## Language, Conflict, and Peace-building

---

### Background on the Language, Conflict and Peace-building Initiative

At this time in our country and the larger global community, many of us are struggling to find ways to talk to each other about conflicts, whether they are occurring in the United States or abroad. As a result, our conversations often leave us polarized and divided, while doing little or nothing to help resolve the critical issues at the heart of those conflicts.

The LSA Executive Committee has held a series of discussions regarding how the insights, tools, and empirical findings of linguistics could be highlighted, organized, and engaged so as to illuminate the ways language can be used as a tool to provoke conflict and drive polarization, or to build bridges for peace-building.

Four ideas have emerged from those conversations that could perhaps build on one another. They are broad ideas for a collaborative, inclusive approach that begins by working locally, with awareness of the global context. The Executive Committee has empowered the Secretariat and President to continue developing these ideas, identifying individuals interested in contributing to them, and working toward implementation of those ideas that emerge as viable.



# References

- Allen, R. (2022) The Important Difference between Emotions and Feelings, Psychology Today [online article]. Accessed 10/27/2025 from: <https://www.psychologytoday.com/us/blog/the-pleasure-is-all-yours/202202/the-important-difference-between-emotions-and-feelings>
- Biadys, Fadi, Soltau, Hagen, Mangu, Lidia, Navratil, Jiri, and Hirschberg, Julia (2010) Discriminative Phonotactics for Dialect Recognition Using Context-Dependent Phone Classifiers. Odyssey 2010, Jun 10-Jul 1
- Bender, Emily (2022) "The place of linguistics and symbolic structures", NAACL 2022 Plenary Panel (with Chitta Baral, Emily Bender, Dilek Hakkani-Tur, Christopher Manning, Dan Roth). doi: 10.48448/0ep5-cm79
- Hui, Jonathan (2019) "Speech Recognition –GMM, HMM", The Medium.com accessed online 1/11/2021
- diPaolo, M. and M. Yaeger-Dror (2011) Sociophonetics: a Student's Guide. Routledge.
- Dorn, R., 2019. Dialect-specific models for Automatic Speech Recognition of African American Vernacular English, in: Proceedings of the Student Research Workshop associated with RANLP. 16-20.
- Fridland, V., T. Kendall, A.B. Wassink, B. Evans & L. Hall-Lew (2016, 2017, 2020) Speech in the Western States, 3 vols. Publications of the American Dialect Society 101-103. Durham, NC: Duke University Press
- Glaser, Barney G. and Anselm Strauss. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago, IL: Aldine Publishing Co, 1967.
- Harrell, S. P. (2000). A multidimensional conceptualization of racism-related stress: Implications for the well-being of people of color. American Journal of Orthopsychiatry, 70, 42-57.
- Harwell, D. B. Mayes, M. Walls, S. Hashemi, The accent gap. The Washington Post, 19 July 2018. <https://www.washingtonpost.com/graphics/2018/business/alexa-does-not-understand-your-accent/>. Accessed 28 February 2020.
- Jain, Abhinav, Singh, Vishwanath P. and Rath, Shakti. (2019). A Multi-Accent Acoustic Model using Mixture of Experts for Speech Recognition. Presented at Interspeech 2019, Graz Austria.
- Jain, Mahaveer (2022) "In Depth Explanation Of RNN-based Automatic Speech Recognition Systems (ASR)". <https://www.youtube.com/watch?v=n-4osSfmop0&list=PLapTxW9vHtzqMo6gS5zn5YCNrz2xAm9W&index=4>
- Jurafsky, D & Martin, J.H. 2023. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Sections 9.1, 9.2, 9.4.
- Koenecke, Allison, Nam, Andrew, Lake, Emily, Nudell, Joe, Quartey, Minnie, Mengesha, Zion, Toups, Connor, Rickford, John R., Jurafsky, Dan, & Goel, Sharad (2020) Racial disparities in automated speech recognition, Proc. of the National Academy of Sciences, 117(14), April 7: 7684-7689.
- Papstratis, Ilias (2021) Speech Recognition: a review of the different deep learning approaches. AI Summer accessed from: <https://theaisummer.com/speech-recognition/#deep-neural-networks-for-asr>
- Pierce, C. M. (1995). Stress analogs of racism and sexism: Terrorism, torture, and disaster. In C. V. Willie, P. P. Reiker, B. M. Kramer, & B. S. Brown (Eds.), Mental health, racism, and sexism (pp. 277-293). Pittsburgh: University of Pittsburgh Press.
- Raji, D. [@rajiinio]. 2021, March 27. "1. Bias can start anywhere in the system - pre-processing, post-processing, with task design, with modeling choices, etc., in addition to issues with the data. The system arrives as the result of a lot of decisions, and any of those decisions can result in a biased outcome." [Tweet]. <https://twitter.com/rajiinio/status/1375957288276611075> (accessed 2022, January 31).
- Squizzero, R., Horst, M. Wassink, A.B., Panicacci, A., Jensen, M. Moroz, A., Conrad, K., & Bender, E. (2021) Collecting and using race and ethnicity information in linguistic studies, University of Washington Working Papers in Linguistics.
- Tatman, R., & Kasten, C. 2017. Effects of Talker Dialect, Gender & Race on Accuracy of Bing Speech and YouTube Automatic Captions. INTERSPEECH 2017, 934-938
- Wassink, A.B. (2020) "Where Sociolinguistics and Speech Science Meet: The physiological and acoustic consequences of underbite in a multilectal speaker of African-American English." In Buchstaller, I. and Blake, R. The Routledge Companion to the Work of John Rickford. New York: Routledge.
- Wassink, Alicia (2017) The Vowels of Washington State In Speech in the Western States: vol. 1: the coastal states (Fridland, V. Kendall, T. Wassink, A.B. and Evans, B., eds.). Publications of the American Dialect Society. Duke University Press.
- Wassink, Alicia, Gansen, Cady, and Bartholomew, Isabel (2022) Uneven success: automatic speech recognition and ethnicity-related dialects, Speech Communication, 140:50-70.
- Wolfram, W., & Schilling-Estes, N. (1998) American English, ch. 4. Dialects in the US: past, present and future. Oxford, UK: Blackwell, pp. 90-123.
- Xiong, W., Wu, L., Alleva, F., Droppo, J., Huang, X., Stolcke, A., 2017. The Microsoft 2017 Conversational Speech Recognition System. Microsoft AI and Research Technical Report MSR-TR-2017-39.
- Yuan, J. and M. Liberman (2008) Penn Phonetics Lab Forced Aligner (P2FA). Speaker identification on the SCOTUS corpus. Proceedings of Acoustics.