

Socially Robust AI

Maria Teleki

PhD Candidate, Computer Science
Texas A&M University



Introduction

Research

Teaching

Collaborations

I build methods, benchmarks, and audits that **improve AI robustness in socially complex situations.**

My research spans Speech AI and Computational Social Science.

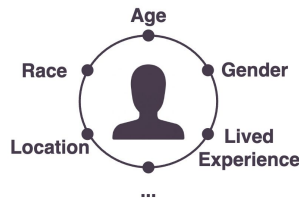
Maria Teleki

Socially Robust AI



Is AI robust to how people naturally speak?

TACL '26, ICASSP '26
INTERSPEECH '25/'24,
LREC-COLING '24 x2



Is AI robust to each person's uniqueness?

ICWSM '25, INTERSPEECH '26
ACL '26, EMNLP Findings '23/'25,
IUI '26, CHI '26



Is AI robust within institutions?

FAcT '26
L@S '26



Introduction

Research

Teaching

Collaborations

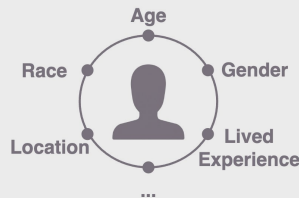
Maria Teleki

Socially Robust AI



Is AI robust to how people naturally speak?

TACL '26, ICASSP '26
INTERSPEECH '25/'24,
LREC-COLING '24 x2



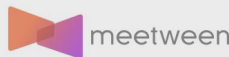
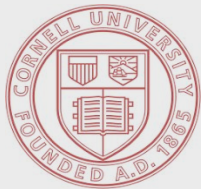
Is AI robust to each person's uniqueness?

ICWSM '25, INTERSPEECH '26
ACL '26, EMNLP Findings '23/'25,
IUI '26, CHI '26



Is AI robust within institutions?

FACCT '26
L@S '26



Introduction

Research

Teaching

Collaborations

Speech AI is everywhere

Set a timer for seven, uh, minutes.



Smart Speaker

Find me the nearest **gas station** — no, sorry, **grocery store**.



Smart Phone

Where is my wallet, **you know**, the black one I always carry?



Smart Glasses

Voice Assistants



Siri Alexa Claude, ChatGPT, Gemini
all have Voice Mode

Healthcare

Transcription
Telehealth
Automated triage
Biomarkers for dementia, depression, ALS, and more
Check-in systems
Record keeping (SOAP)
Patient scheduling

Education

Online courses
Tutoring & coaching systems

Legal

Court transcription

Entertainment

Dubbing
Music generation

Accessibility

Live translation (incl. code-switching)
Expressive captions
Illiterate users

Business

Meeting summarization, proactive agents
Voice-to-code programming
In-car voice assistants
Voice banking (finance)

Introduction

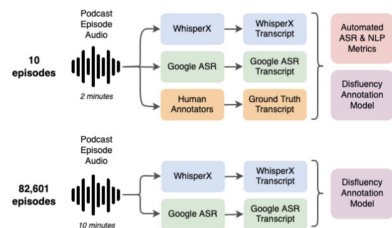
Research

Teaching

Collaborations

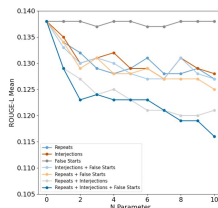
My research established...

Disfluencies are OOD



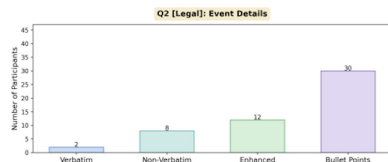
INTERSPEECH '24

Disfluencies harm task performance in...



summarization
LREC-COLING '24

Disfluency removal or preservation depends on task



In Prep for TOCHI

rec. systems.
INTERSPEECH '25

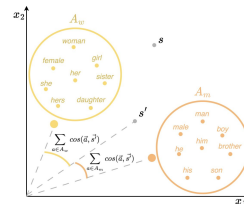
A vegetarian one without any peanuts in it would be great.

Preferences = {'dishName': 'cake',
'occasion': 'valentine's day', 'dietaryFilters':
['vegetarian', 'peanut free']}



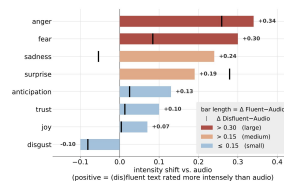
translation
TACL '26

Disfluencies carry identity information, e.g.,



gender
ICWSM '25

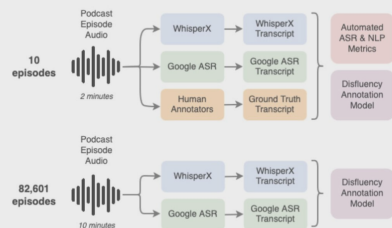
Disfluencies carry emotional information



TACL '26

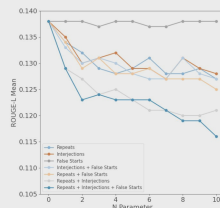
My research established...

Disfluencies are OOD



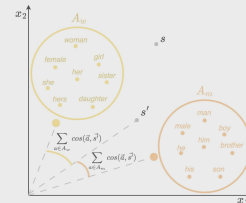
INTERSPEECH '24

Disfluencies harm task performance in...



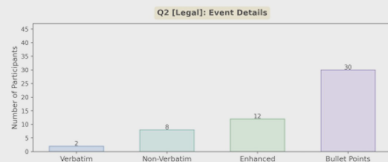
summarization
LREC-COLING '24

Disfluencies carry identity information, e.g.,



gender
ICWSM '25

Disfluency removal or preservation depends on task



In Prep for TOCHI

rec. systems.
INTERSPEECH '25

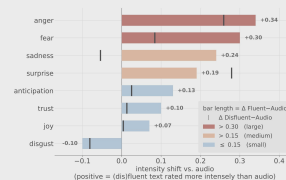
A vegetarian one without any peanuts in it would be great.

Preferences = {'dishName': ['cake'],
'occasion': ['valentine's day'], 'dietaryFilters':
['vegetarian', 'peanut free']}



translation
TACL '26

Disfluencies carry emotional information

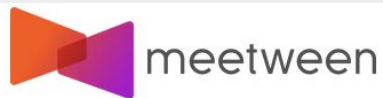


TACL '26

Disfluency-Aware Translation



Applications in healthcare, business,
disaster response, and more



The disfluencies could mean...

- trying to remember what their dose is
- uncertainty about the dose
- anxiety about talking to a doctor (power imbalance)
- searching for words in a second language
- a speech or language condition (stuttering, aphasia)



How do speech translation systems translate disfluencies? **TACL '26**

Uh-Mazing benchmark: 80 Switchboard utterances x 8 target languages = 640 human translations with annotated disfluency spans

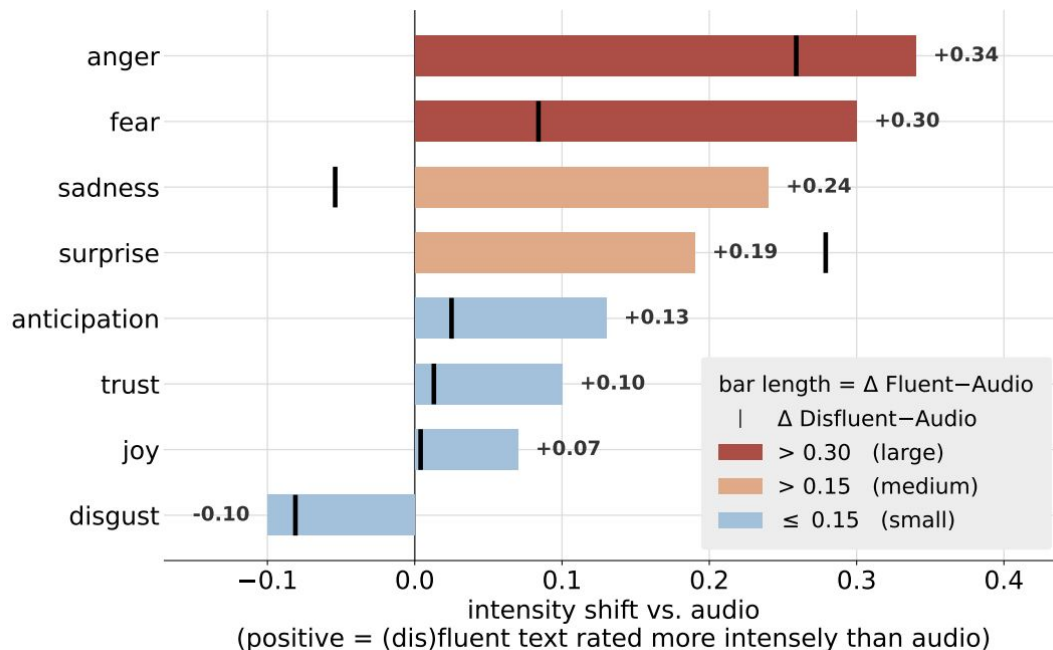
- **Human study** showing disfluencies carry emotional information
- **Taxonomy of failure modes**, including disfluency-specific failure modes
- **Inference-time mitigation strategies** (beam search, in-context learning)



How do speech translation systems translate disfluencies? *TACL '26*

We conducted a human study showing disfluencies carry emotional information:

Key Finding:
Removing disfluencies **increases** the perceived intensity of negative emotions





How do speech translation systems translate disfluencies? **TACL '26**

		by language								by model					
		AR	CS	DE	ES	FR	HI	IT	ZH	total	canary	gemini	tower	phi	total
General Errors	Incorrect language ★	203	93	33	49	76	42	141	84	721	117	104	207	293	721
	Missing content ★	67	34	52	58	44	23	193	50	521	133	151	84	153	521
	Extra content	13	10	30	38	17	11	70	30	219	47	56	55	61	219
	Looping	19	23	6	6	4	16	1	7	82	2	4	15	61	82
Disfluency Errors	Deletion ★	11	68	144	208	184	88	238	157	1098	353	276	165	304	1098
	Extra disfluency	55	5	38	27	8	19	61	9	222	32	36	82	72	222
	Type mismatch	19	27	6	17	11	35	101	8	224	50	49	67	58	224
	Other error		1	2	2					5		3	2		5
Other Errors	Uncategorized error	9	3	17	10	20		1	2	62	14	12	17	19	62
		396	264	328	415	364	234	806	347	3154	748	691	694	1021	3154

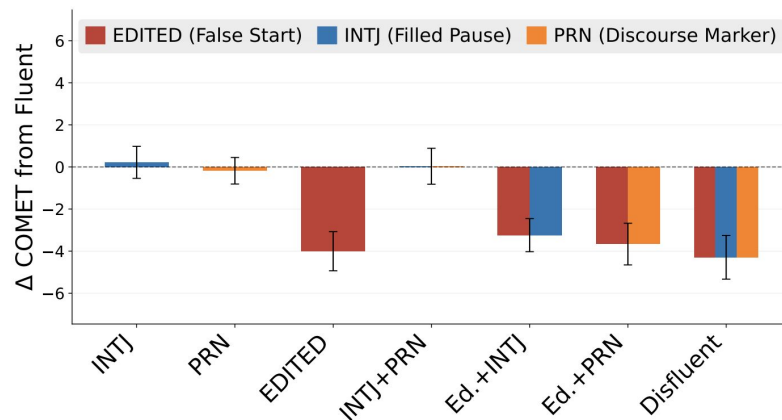
We identify a **taxonomy of failure modes**.



How do speech translation systems translate disfluencies? **TACL '26**

We make a **linguistic diagnosis** of what drives the failures:

EDITED disfluencies **drive the drop in performance**; some can be **recovered**, detailed next...

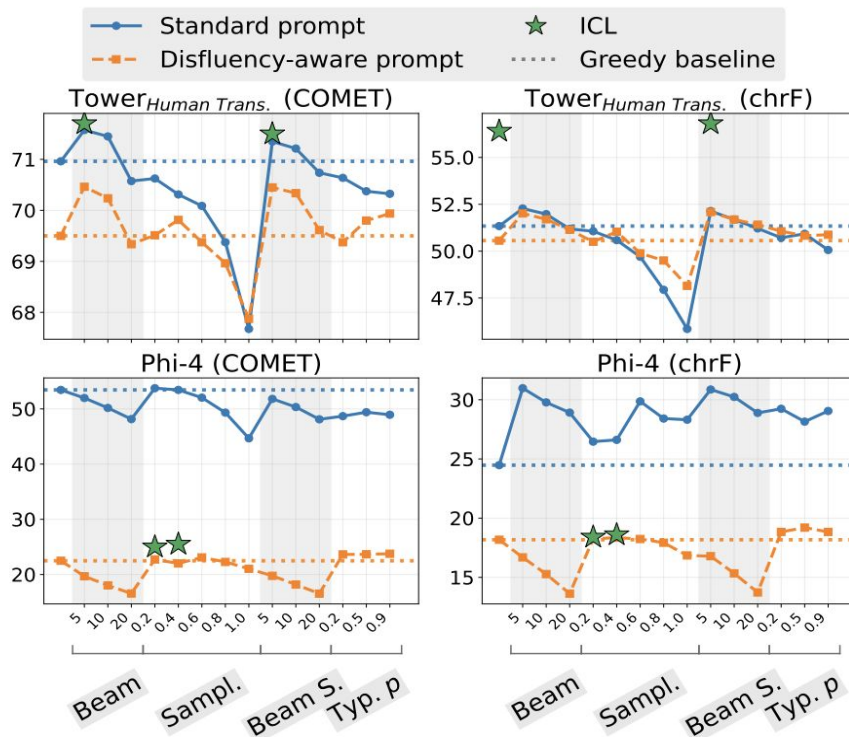




How do speech translation systems translate disfluencies? *TACL '26*

We contribute **inference-time mitigation strategies** (beam search, in-context learning) to recover some of the performance loss:

Decoding methods improve overall quality (COMET), while **ICL improves disfluency preservation** (chrF)



1

Modeling Generation Axes: Generation Model, Judge Model, Temperature, Reasoning Trace Levels (Global, Per-Turn, Per-Burst), ...

Sociodemographic Generation Axes: Age, Race, Gender, Gender Presentation, ...

Narrative Generation Axis: Plot Shape (Vonnegut)

2

Isolate a specific interruption
for judgement

A: The committee report is basically ready to—

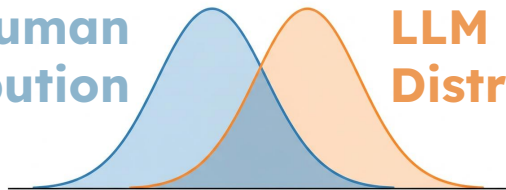
B: [=Right, but we haven't looked at the ↑ budget numbers—=]

A: [=—and we can file it by ↓ Monday=] if you just let me finish.

B: Oh, go ahead. (0.4) I'm ::: listening ::: ↑ really.

3

Human
Distribution



LLM Judge
Distribution

+ compare
free-form
reasoning traces

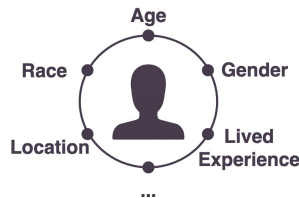
Maria Teleki

Socially Robust AI



Is AI robust to how people naturally speak?

TACL '26, ICASSP '26
INTERSPEECH '25/'24,
LREC-COLING '24 x2



Is AI robust to each person's uniqueness?

ICWSM '25, INTERSPEECH '26
ACL '26, EMNLP Findings '23/'25,
IUI '26, CHI '26



Is AI robust within institutions?

FAcT '26
L@S '26



Introduction

Research

Teaching

Collaborations

Outreach

Invited Talks

Academic: George Mason University, King's College London (upcoming), Texas Tech University (Vet Med)

Industry: JLL and IPT Global in Houston, Revionics in Austin

Students: Club of Aggie Female Engineers, Sikorski Horizon (upcoming)



→ A Cross-Community Agenda for Speech AI out of CHI '26 workshop; 23 authors spanning 18 institutions

→ Working with INTERSPEECH to open a new track for cross-community speech AI work

Teaching and Mentoring

CSCE 676: Data Mining & Analysis (*Graduate, 125 students*)

Course Developer, TA – helped redesign the course for the AI era: e.g., students use coding agents, then validate the agent's output vs. established libraries, students practice answering interview questions

Guest Lecture – SpeechLLMs

CSCE 670: Information Storage & Retrieval (*Graduate, 125 students*)

Guest Lectures – IR Evaluation and Learning to Rank

CSCE 289: GenAI for Engineers (*Undergraduate, 30 students*)

Course Developer – course builds from foundational concepts (via hands-on API and pipeline work) to critical evaluation of model outputs

Professional Development Mastery Certificate in Instruction & Assessment, CIRT
Multi-semester program in evidence-based teaching, including syllabus design, learning assessment, and inclusive pedagogy

Teaching and Mentoring



12 mentees (4 MS, 8 UG)

4 first-author papers

- CHI '26 Poster
- 2 x ICWSM '26 Workshop Papers
- LREC-COLING '24

Internships & jobs at
NVIDIA, Amazon, Meta,
Google, Samsung, Target

Multiple students
applying to MS and PhD
programs – 1 currently at
Cornell Tech MS

Introduction

Research

Teaching

Collaborations

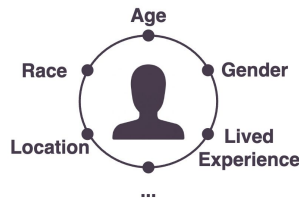
Maria Teleki

Socially Robust AI



Is AI robust to how people naturally speak?

TACL '26, ICASSP '26
INTERSPEECH '25/'24,
LREC-COLING '24 x2



Is AI robust to each person's uniqueness?

ICWSM '25, INTERSPEECH '26
ACL '26, EMNLP Findings '23/'25,
IUI '26, CHI '26



Is AI robust within institutions?

FAcT '26
L@S '26



Introduction

Research

Teaching

Collaborations