

Current and Potential Future Uses of AI in Speech and Language Therapy

Mark Hasegawa-Johnson
October and November 2023

Outline

- Automatic diagnosis and analysis of voice, speech, and sleep
 - Parkinson's: Research results and remaining hurdles to clinical deployment
 - Developmental language disorder: Potential for early screening
 - Autism: Visualizations to assist clinician-parent communication
 - Infant sleep disorders: Potential for low-cost objective transcripts
 - Dysarthria intelligibility: Research results, hurdles to clinical deployment



https://commons.wikimedia.org/wiki/File:Doctor_and_patient,_1509_Wellcome_L0011744.jpg



https://commons.wikimedia.org/wiki/File:Women_practice_voting_in_Dayton_Oct._27,_1920.jpg

- Spoken language access to government, education, and employment
 - Motivation
 - Research results using the UA-Speech corpus
 - Advances since 2008 in speech technology state of the art
 - Speech Accessibility Project
 - Relevance to Speech-Language Therapy

The stages of life

- Parkinson's
 - (typical age of diagnosis: 40-80y)
- Developmental Language Disorder
 - (typical aod: 3-6y)
- Autism
 - (typical aod: 1-3y)
- Infant Sleep Disorders
 - (typical aod: 0-1y)
- Cerebral Palsy
 - (typical aod: 0-1y)



The Stages of Life, Caspar David Friedrich,
https://en.wikipedia.org/wiki/The_Stages_of_Life#/media/File:Caspar_David_Friedrich_013.jpg

Living well with Parkinson's: Medical advances are only useful if used

Neurology®



[Neurology](#). 2018 Nov 27; 91(22): e2045–e2056.

doi: [10.1212/WNL.00000000000006576](https://doi.org/10.1212/WNL.00000000000006576)

PMCID: PMC6282235 | PMID: [30381367](https://pubmed.ncbi.nlm.nih.gov/30381367/)

Early predictors of mortality in parkinsonism and Parkinson disease

A population-based study

[David Bäckström](#), MD, [✉][Gabriel Granåsen](#), MSc, [Magdalena Eriksson Domellöf](#), PhD, [Jan Linder](#), MD, PhD, [Susanna Jakobson Mo](#), MD, PhD, [Katrine Riklund](#), MD, PhD, [Henrik Zetterberg](#), MD, PhD, [Kaj Blennow](#), MD, PhD, and [Lars Forsgren](#), MD, PhD

- Recent studies suggest that people with Parkinson's do not die earlier than people without Parkinson's unless the disease causes cognitive or motor impairments that limit lifespan.
- Treatments minimizing cognitive and motor impairments can provide a normal lifespan.
- Many treatments are only possible after diagnosis.

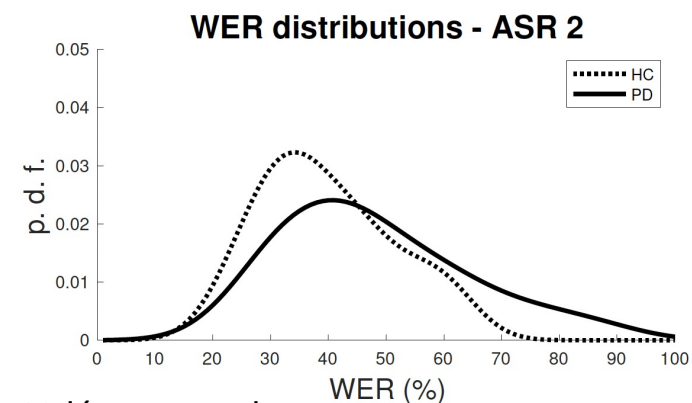
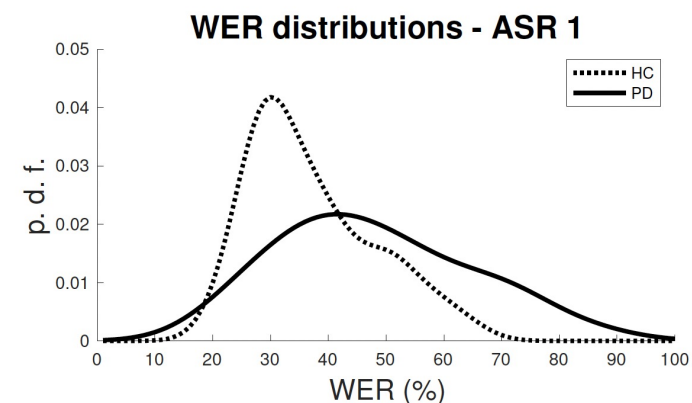


FIGHTING BACK AGAINST PARKINSON'S



Parkinson's early diagnosis from speech

- Trained listeners:
 - Human: "I can tell immediately if someone has Parkinson's"
 - Machine: "...word error rate is 27% higher in speakers with PD than in control speakers..."
- Untrained listeners:
 - Human: "They don't sound any different from anybody else"
 - Machine: "Our rating system says these people have normal intelligibility; I don't think our speech recognizers would have any trouble transcribing them"
- Can we use training to automatically diagnose Parkinson's at an early stage?



Moro-Velázquez et al.,
<http://dx.doi.org/10.21437/Interspeech.2019-2993>

Parkinson's early diagnosis from speech

- PD can be detected with 79-94% accuracy, depending on corpus and speech materials, with speech dynamics being especially useful features (Moro-Velázquez et al., [10.1109/ICASSP40776.2020.9053770](https://doi.org/10.1109/ICASSP40776.2020.9053770)).
- Detection accuracy is worse over an analog telephone channel (Fraile et al., [10.21437/Interspeech.2007-391](https://doi.org/10.21437/Interspeech.2007-391)), but is not harmed by digital compression at a sufficiently high data rate (Sáenz-Lechón et al., [10.1109/TBME.2008.923769](https://doi.org/10.1109/TBME.2008.923769)), so telemedicine is possible.

Table 1: Best cross-validation results for each experiment as a function of the employed corpus and speech task. Sust. v. stands for sustained vowel, and Exp. for Experimental set.

Corpus	Speech task	Exp.	Accuracy \pm CI	AUC	Sens.	Spec.	
GITA	TDU	1	80 \pm 8	0.85	0.82	0.78	
		2	81 \pm 8	0.88	0.84	0.78	
		3	85 \pm 7	0.91	0.82	0.88	
		4	85 \pm 7	0.89	0.84	0.86	
	DDK	1	81 \pm 8	0.88	0.82	0.8	
		2	79 \pm 8	0.86	0.86	0.72	
		3	83 \pm 7	0.89	0.86	0.8	
		4	83 \pm 7	0.88	0.86	0.8	
	Monol.	1	80 \pm 8	0.88	0.76	0.84	
		2	78 \pm 8	0.84	0.73	0.82	
		3	82 \pm 8	0.89	0.8	0.84	
		4	80 \pm 8	0.86	0.76	0.84	
	Sust. v.	5	71 \pm 9	0.8	0.72	0.7	
	TDU + Sust. v.	5	85 \pm 7	0.91	0.82	0.88	
	Neurovoz	TDU	1	86 \pm 8	0.93	0.87	0.84
			2	81 \pm 9	0.87	0.83	0.78
3			89 \pm 7	0.93	0.87	0.91	
4			89 \pm 7	0.93	0.91	0.84	
DDK		1	79 \pm 9	0.85	0.87	0.65	
		2	79 \pm 8	0.86	0.86	0.72	
		3	86 \pm 8	0.88	0.89	0.81	
		4	83 \pm 7	0.88	0.86	0.8	
Monol.		1	79 \pm 12	0.81	0.59	0.9	
		2	66 \pm 14	0.67	0.35	0.83	
		3	77 \pm 12	0.79	0.53	0.9	
		4	79 \pm 12	0.9	0.47	0.97	
Sust. v.		5	64 \pm 10	0.68	0.75	0.48	
TDU + Sust. v.		5	87 \pm 7	0.94	0.85	0.91	
CzechPD		DDK	1	88 \pm 1	0.94	0.85	0.93
			2	94 \pm 1	0.97	0.9	1
	3		94 \pm 1	0.98	0.9	1	
	4		94 \pm 1	0.99	0.9	1	

Remaining obstacles to clinical deployment

Moro-Velázquez et al., <https://doi.org/10.1016/j.bspc.2021.102418>

- Patient privacy laws \Rightarrow datasets with confirmed diagnoses are very small, large datasets only contain proxies like self-diagnosis
 - Diagnosis accuracy for any one patient varies significantly depending on exactly which patients are in the training set.
- The curse of small datasets can be overcome by pooling datasets, but:
 - The population with Parkinson's is older than the population without.
 - Automatic diagnosis algorithms learn to "diagnose" the microphone.
- A clinical trial would solve the problem, but:
 - Nobody has yet successfully made the case that a voice-based test provides benefit that is not available in standard eldercare clinical visits, especially if accuracy is less than 94%.
- Larger datasets, made available to researchers, might increase accuracy.
 - NIH Bridge2AI Program has funded "[Voice as a Biomarker for Health](#)" project, which is surveying available datasets.

The stages of life

- Parkinson's
 - (typical age of diagnosis: 40-80y)
- Developmental Language Disorder
 - (typical aod: 3-6y)
- Autism
 - (typical aod: 1-3y)
- Infant Sleep Disorders
 - (typical aod: 0-1y)
- Cerebral Palsy
 - (typical aod: 0-1y)







The Stages of Life, Caspar David Friedrich,
https://en.wikipedia.org/wiki/The_Stages_of_Life#/media/File:Caspar_David_Friedrich_013.jpg

Developmental language disorder (DLD)

- In the United States, more than one-half of children under the 13 categories in the Individuals with Disabilities Education Act (IDEA) need speech and language services (3.4m children: <https://sites.ed.gov/idea/>)
- Identifying children in need of speech and language therapy is important, because early intervention improves the child's academic and socio-economic prospects (Kaiser & Roberts, <https://doi.org/10.1177/1053815111429968>)
- Identifying children in need of speech and language therapy is challenging because more than half of US school districts report teaching vacancies, among which speech and language pathology is one of the most frequently reported (Squires, <https://eric.ed.gov/?id=EJ1135588>)

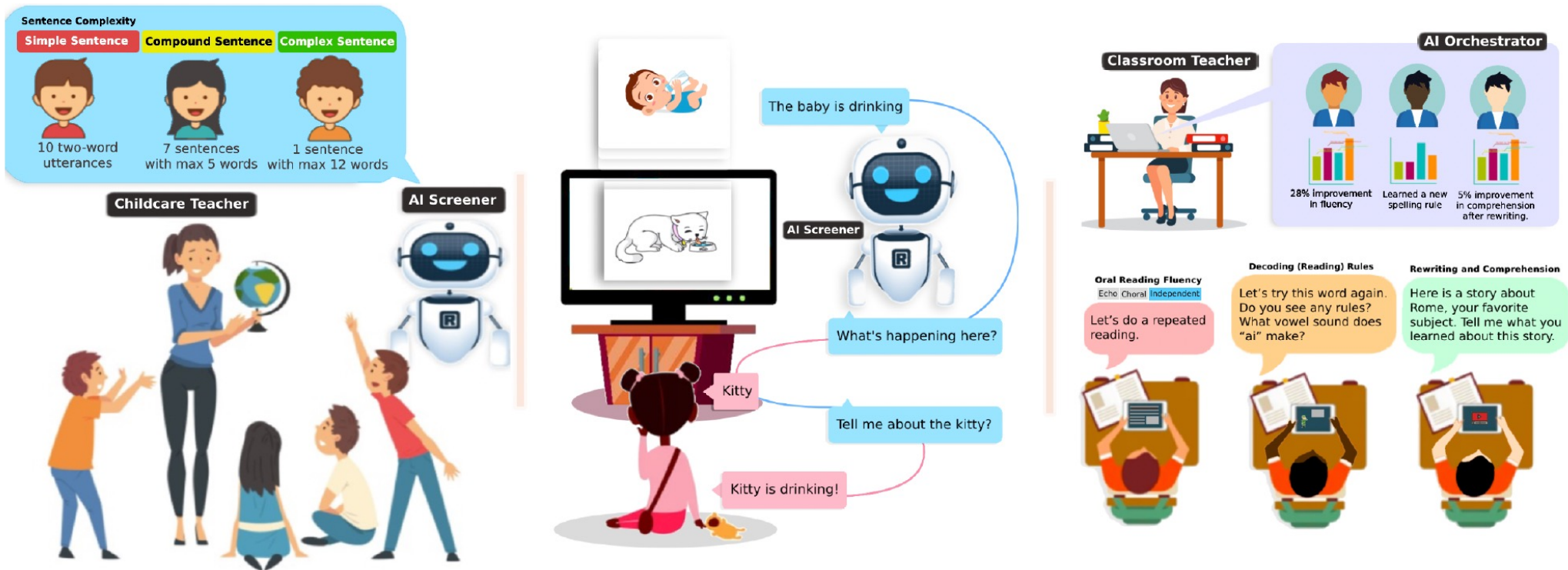
DLD: Screening and assessment

- Comprehensive language assessment is the gold standard but suffers from child non-compliance and visit scheduling problems.
- Parental report (e.g., CDI) is widely used but suffers from variable parent attentiveness/compliance.
- Language sample assessment is variable and time-consuming.
- Telehealth assessment may be performed using structured elicitation protocols such as the sentence diversity priming task (right).

<p>Adult Prime</p>  <p>Point to the picture and say: <i>The baby is drinking</i></p>	
<p>Child Target</p>  <p>Point to the picture and say: <i>What's happening here?</i></p> <p>If the child doesn't respond: Point again and say: <i>Tell me about the kitty!</i></p>	

Krok et al., <https://doi.org/10.1097/TLD.0000000000000280>

Goals of a current research program: AI screening and therapy support for children with developmental language disorder



Driving use cases of the [NSF AI Institute for Exceptional Education](#): Govindaraju, Xiong, Setlur et al., 2022

The stages of life

- Parkinson's
 - (typical age of diagnosis: 40-80y)
- Developmental Language Disorder
 - (typical aod: 3-6y)
- Autism
 - (typical aod: 1-3y)
- Infant Sleep Disorders
 - (typical aod: 0-1y)
- Cerebral Palsy
 - (typical aod: 0-1y)



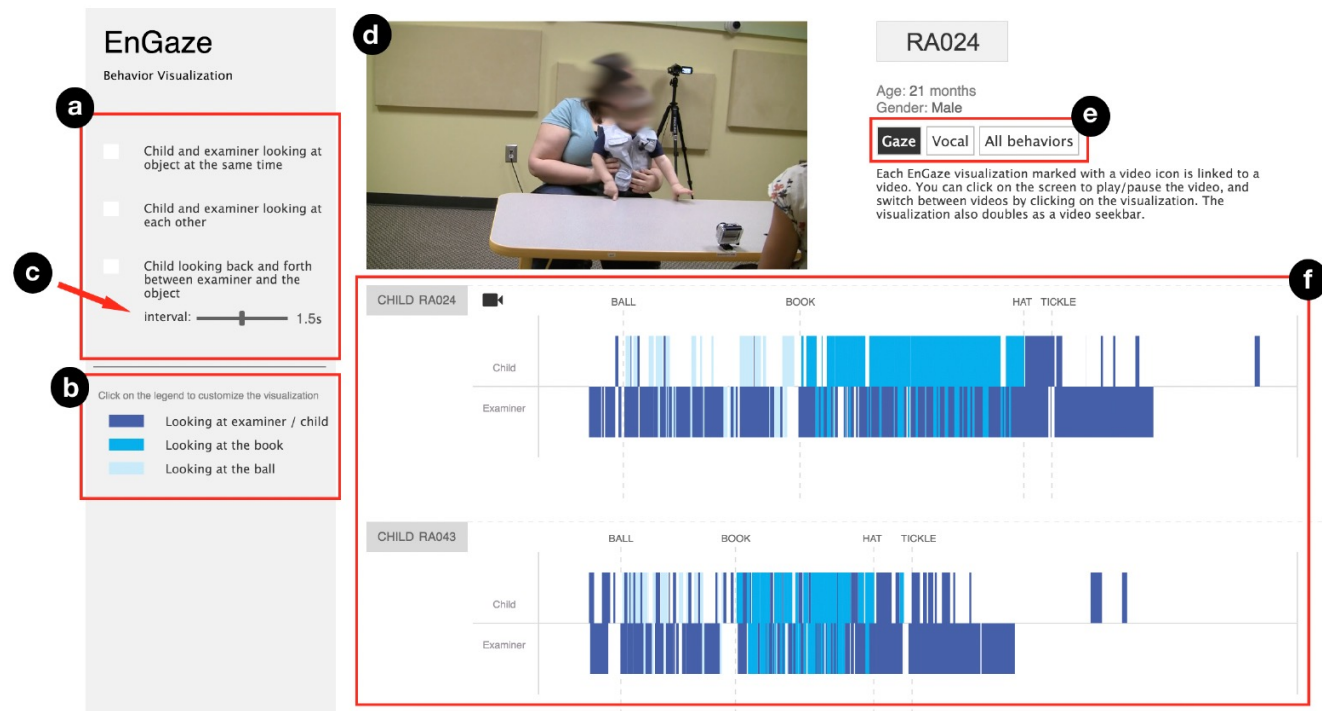
The Stages of Life, Caspar David Friedrich,
https://en.wikipedia.org/wiki/The_Stages_of_Life#/media/File:Caspar_David_Friedrich_013.jpg

Autism spectrum disorder (ASD): Assessment

- There is no cure for autism, but recent evidence suggests that early diagnosis can lead to significantly improved life outcomes
 - e.g., Elder et al., <https://doi.org/10.2147/PRBM.S117499>
- Validated assessment instruments:
 - In-person assessment, called the “gold standard” (Morrier et al., <https://doi.org/10.1007/s10803-023-06116-1>)
 - Parent questionnaire
 - Telehealth assessment

ASD: Screening prior to assessment

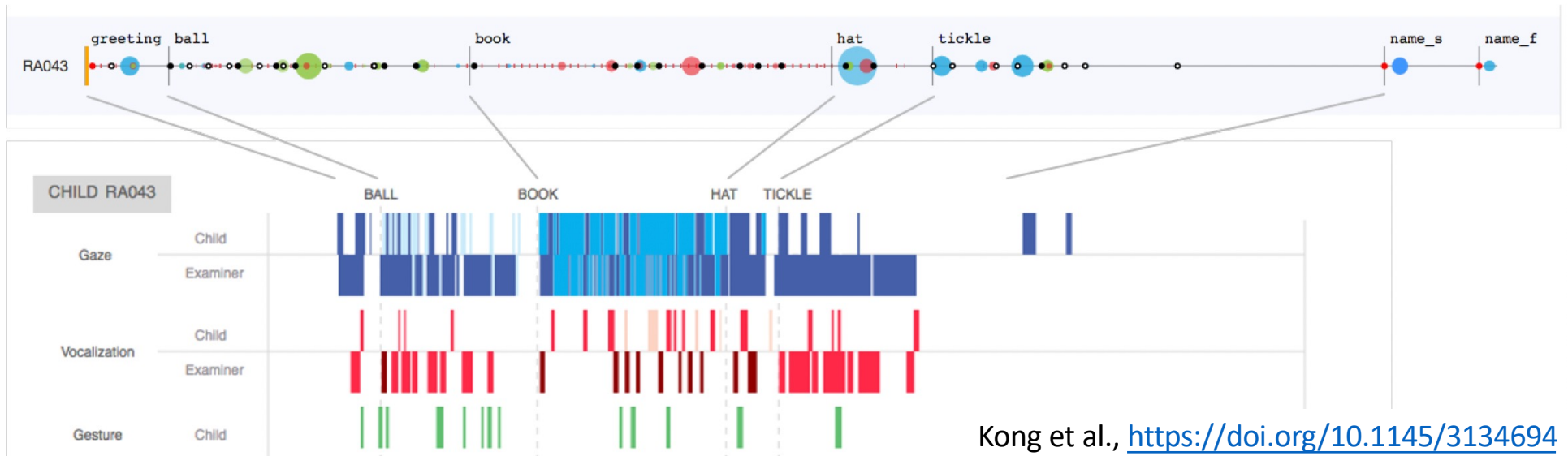
- In addition to parent report, pediatric screening often uses a brief clinician-child interaction to decide whether formal assessment is needed
- Rapid-attention back-and-forth communication ([Rapid ABC](#); Ousley et al., 2009) is a structured 5-minute screening for children aged 15 to 17 months



Kong et al., Proc. ACM HCI 1:59, 2017

ASD: Visualization of screening results

- Screening can identify children in need of further assessment
- Screening is fast, therefore explaining the results to a parent can be challenging
- Visualizations that show gaze, speech and gesture can be used to explain clinician concerns to a parent (Kong et al., <https://doi.org/10.1145/3134694>)
- With two microphones, AI-based automatic detection of speaker turns can be performed with 83% accuracy (Li et al., <https://doi.org/10.21437/Interspeech.2023-460>)



The stages of life

- Parkinson's
 - (typical age of diagnosis: 40-80y)
- Developmental Language Disorder
 - (typical aod: 3-6y)
- Autism
 - (typical aod: 1-3y)
- Infant Sleep Disorders
 - (typical aod: 0-1y)
- Cerebral Palsy
 - (typical aod: 0-1y)



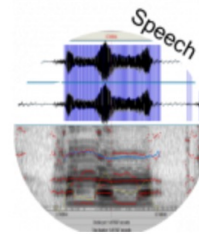
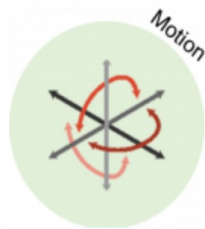
The Stages of Life, Caspar David Friedrich,
https://en.wikipedia.org/wiki/The_Stages_of_Life#/media/File:Caspar_David_Friedrich_013.jpg

Sleep disorders

- Conflicting advice about the best course of action can leave parents uncertain
- Some types of sleep disorder should not be ignored, e.g., in some cases parent presence is a better therapy (France et al., Current Paediatrics 2003)

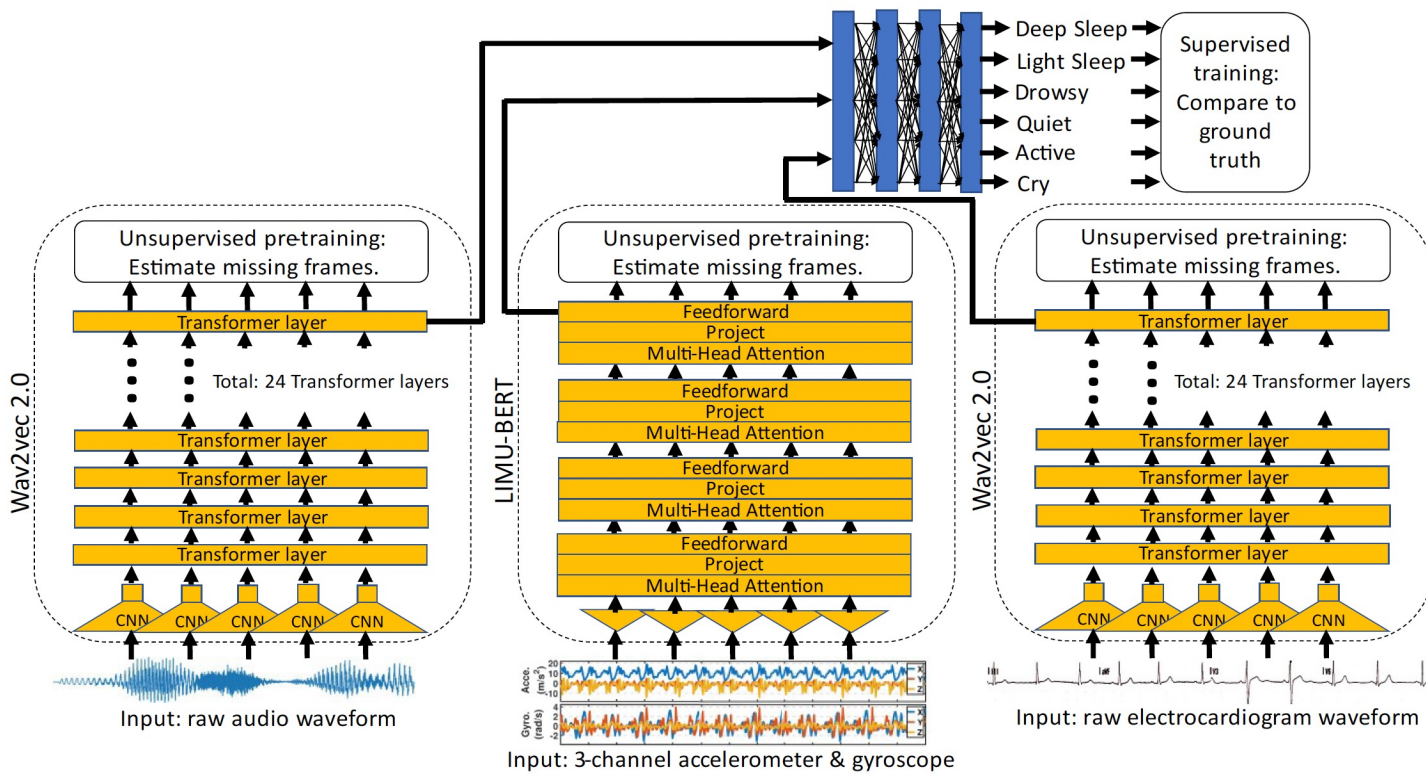
LittleBeats™

- LittleBeats™ (LB) is a multimodal infant-wearable device that simultaneously records, **audio, movement of the child (IMU), and heart-rate variability of the child (ECG)**



<https://littlebeats.hdfs.illinois.edu/>

LittleBeats™ specially-designed shirt



AI monitoring of sleep/wake states using Littlebeats

Automated Assessment of Infant Sleep/Wake States, Physical Activity, and Household Chaos Using a Multimodal Wearable Device and Deep Learning Model: McElwain et al., proposal, 2023

Sleep/Wake Binary Classification	Audio only	ECG only	IMU only	Audio+ECG	Audio+ECG+IMU
Accuracy	88.5%	93.8%	92.5%	92.5%	98.9%
Kappa	0.766	0.873	0.918	0.847	0.977
Precision	0.844	0.913	0.815	0.893	0.995
Recall	0.893	0.942	1.000	0.933	0.978
F Score	0.868	0.928	0.898	0.913	0.986

The stages of life

- Parkinson's
 - (typical age of diagnosis: 40-80y)
- Developmental Language Disorder
 - (typical aod: 3-6y)
- Autism
 - (typical aod: 1-3y)
- Infant Sleep Disorders
 - (typical aod: 0-1y)
- Cerebral Palsy
 - (typical aod: 0-1y)



The Stages of Life, Caspar David Friedrich,
https://en.wikipedia.org/wiki/The_Stages_of_Life#/media/File:Caspar_David_Friedrich_013.jpg

Dysarthria as a symptom of Cerebral Palsy

- The UA-Speech corpus (Kim et al., <https://doi.org/10.21437/Interspeech.2008-480>) contains speech of 17 adults with CP
- Listeners unfamiliar with dysarthria tried to transcribe each isolated word; percentage accuracy is provided as a metric of dysarthria severity

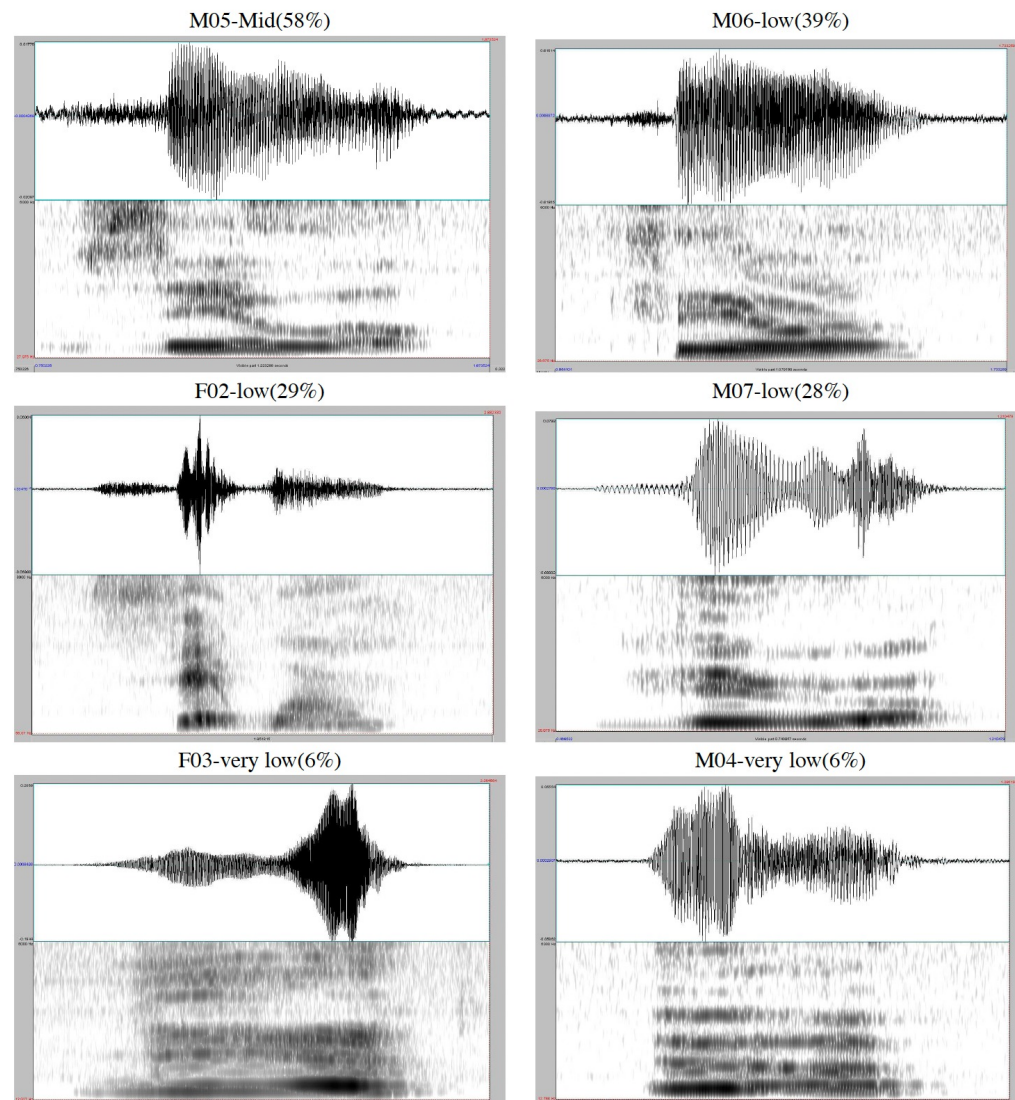


Figure 2: Waveforms and spectrograms of the word 'zero' produced by six different speakers: Speaker code - Intelligibility (%).

Automatic estimation of the intelligibility of dysarthric speech

- Speaker intelligibility can be estimated with $r = 0.98$ by using ASR to transcribe the word, and comparing to what the speaker was trying to say (Tripathi et al., <https://doi.org/10.1109/ICASSP40776.2020.9053339>)
- Correlation coefficients drop a lot when tested on a speaker who was not in the training set (Martinez et al., <https://doi.org/10.1145/2746405>)

Martinez et al., ACM Trans. Accessible Computing, 2015

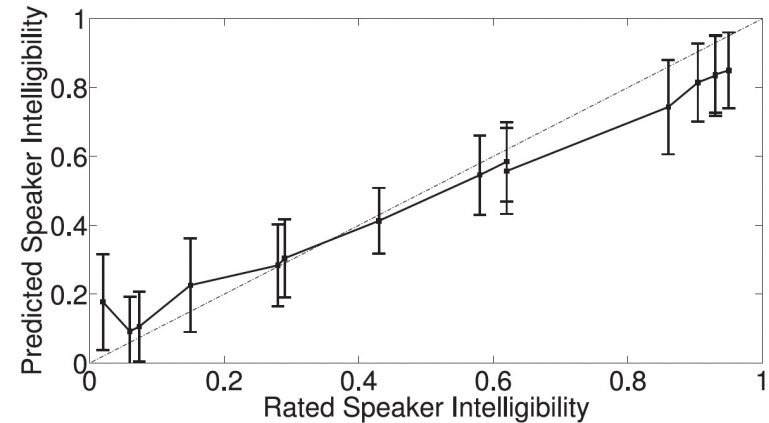


Fig. 2. Mean and standard deviation of intelligibility predictions for each speaker when user data were included in the training dataset (straight) and $x = y$ line (dash-dot).

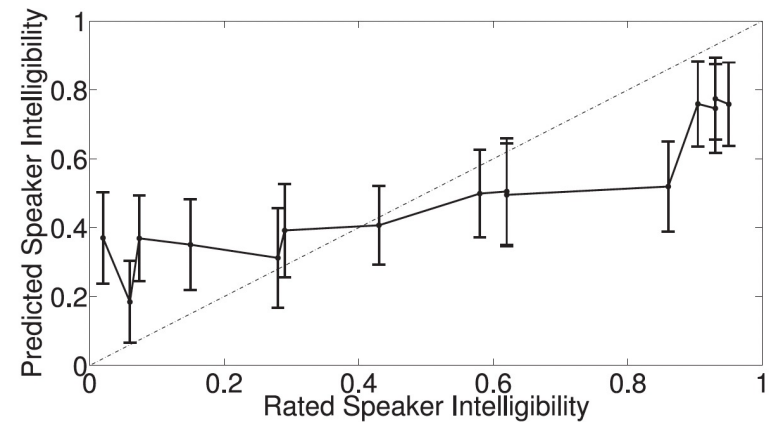


Fig. 3. Mean and standard deviation of intelligibility predictions for each speaker when user data were not included in training dataset (straight) and $x = y$ line (dash-dot).

A few preliminary takeaways

- AI-based screening for dysarthria (e.g., as a symptom of Parkinson's and CP) suffers from out-of-sample generalization error, apparently caused by the small number of speakers in most corpora that are available to researchers
- Providing visualizations and auxiliary information to human clinicians is currently a more promising approach, e.g.,
 - Provide sleep-state timing and durations to screen sleep disorders
 - Provide speech, gaze, and gesture visualizations to screen for ASD
 - (Current research): Provide subject & verb transcripts to screen for DLD

Outline

- Automatic diagnosis and analysis of voice, speech, and sleep
 - Parkinson's: Research results and remaining hurdles to clinical deployment
 - Developmental language disorder: Potential for early screening
 - Autism: Visualizations to assist clinician-parent communication
 - Infant sleep disorders: Potential for low-cost objective transcripts
 - Dysarthria intelligibility: Research results, hurdles to clinical deployment



https://commons.wikimedia.org/wiki/File:Doctor_and_patient,_1509_Wellcome_L0011744.jpg



https://commons.wikimedia.org/wiki/File:Women_practice_voting_in_Dayton_Oct._27,_1920.jpg

- Spoken language access to government, education, and employment
 - Motivation
 - Research results using the UA-Speech corpus
 - Advances since 2008 in speech technology state of the art
 - Speech Accessibility Project
 - Relevance to Speech-Language Therapy

Human rights

Where do rights come from?

“Recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world.”

- Universal Declaration of Human Rights, Preamble

Access is a human right

“Everyone has the right of equal access to public service in his country.”

- UDHR Article 21

“Everyone has the right to work.”

- UDHR Article 23

“Higher education shall be equally accessible to all on the basis of merit.”

- UDHR Article 26

The Speech Accessibility Project Could Open Doors, Literally



A brief biased history of speech technology for people with motor disorders

- 1990: Dragon Dictate allows control of a PC using only speech, “and found acceptance among the disabled” (Maher, 2023)
- 1993: “Speech input for dysarthric users,” (Hwa-Ping Chang, JASA 94:1782)
- 1985-2018: Stephen Hawking uses the “Perfect Paul” speech synthesizer
- 2018: “A phenomenological look at the life hacking-enabled practices of individuals with mobility and dexterity impairments,” Jerry Robinson

[< Back to blog](#)

**Stephen
Hawking's
voice, made
by a man who
lost his own**

Rachel Handley | 15.Jul.2021



Outline

- Automatic diagnosis and analysis of voice, speech, and sleep
 - Parkinson's: Research results and remaining hurdles to clinical deployment
 - Developmental language disorder: Potential for early screening
 - Autism: Visualizations to assist clinician-parent communication
 - Infant sleep disorders: Potential for low-cost objective transcripts
 - Dysarthria intelligibility: Research results, hurdles to clinical deployment



https://commons.wikimedia.org/wiki/File:Doctor_and_patient,_1509_Wellcome_L0011744.jpg



https://commons.wikimedia.org/wiki/File:Women_practice_voting_in_Dayton_Oct._27,_1920.jpg

- Spoken language access to government, education, and employment
 - Motivation
 - Research results using the UA-Speech corpus
 - Advances since 2008 in speech technology state of the art
 - Speech Accessibility Project
 - Relevance to Speech-Language Therapy

UA-Speech

- <https://doi.org/10.21437/Interspeech.2008-480>
- 17 participants with dysarthria (mostly spastic, one mixed) as a symptom of Cerebral Palsy
- 16 control participants, matched for age and gender
- Read isolated words (computer commands, digits, radio alphabet, common words, uncommon words)
- Participants agreed to make their data available to government and academic researchers
- Has been downloaded by hundreds of researchers in ≈ 30 countries

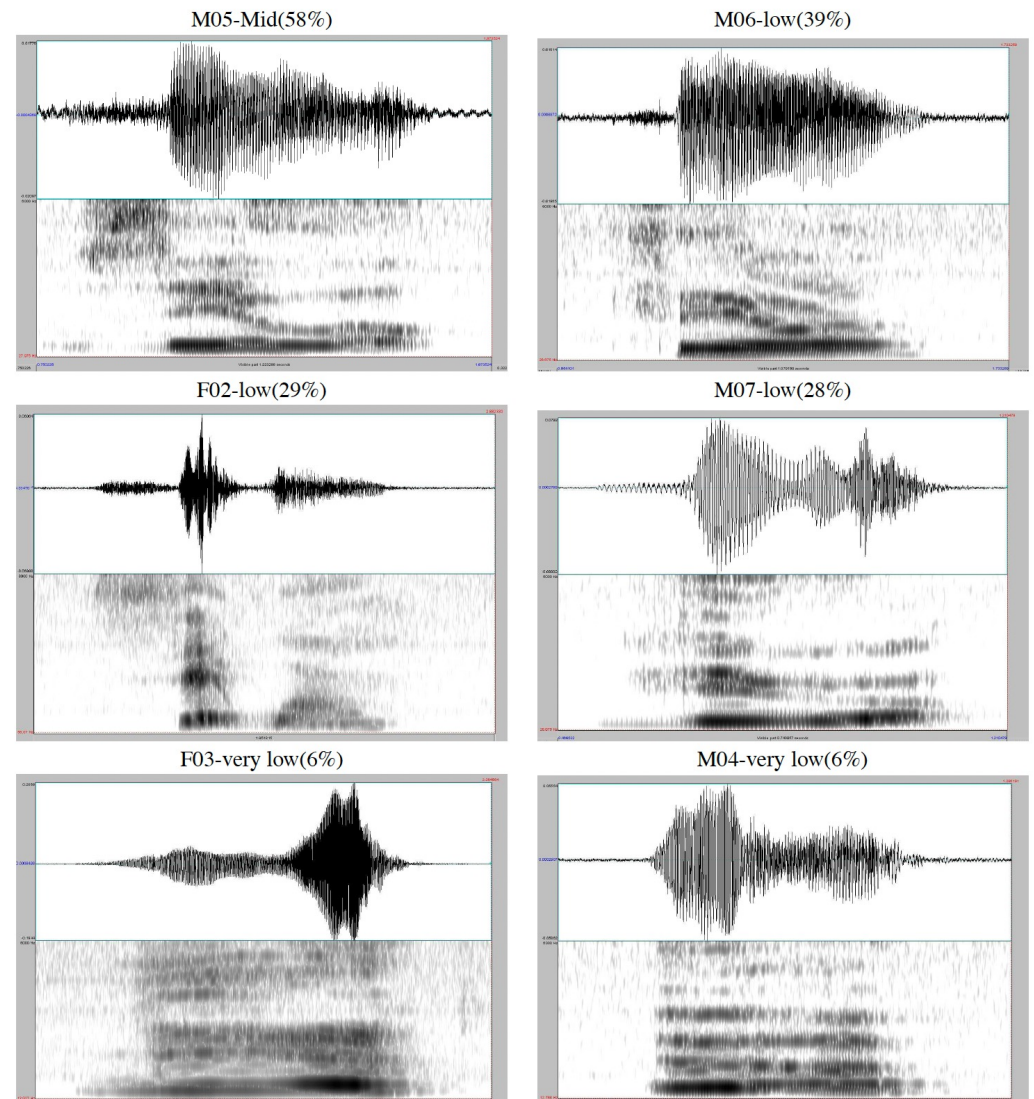
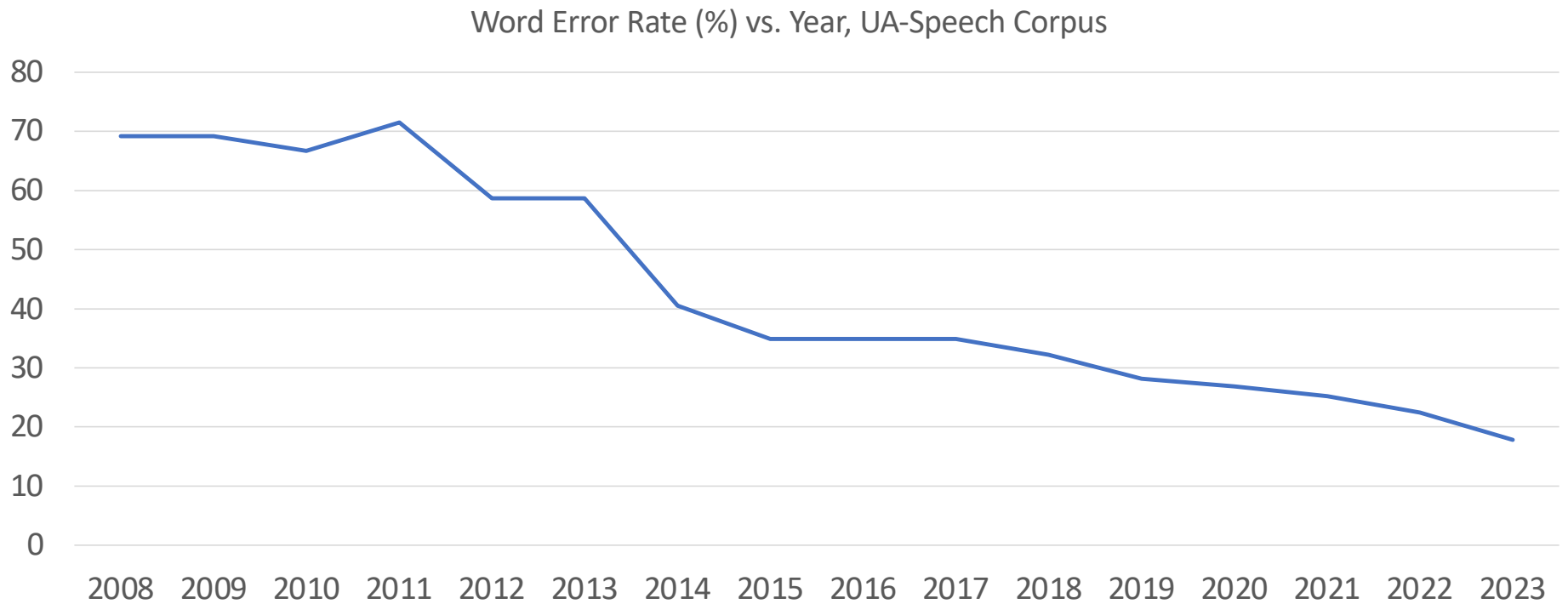


Figure 2: Waveforms and spectrograms of the word 'zero' produced by six different speakers: Speaker code - Intelligibility (%).

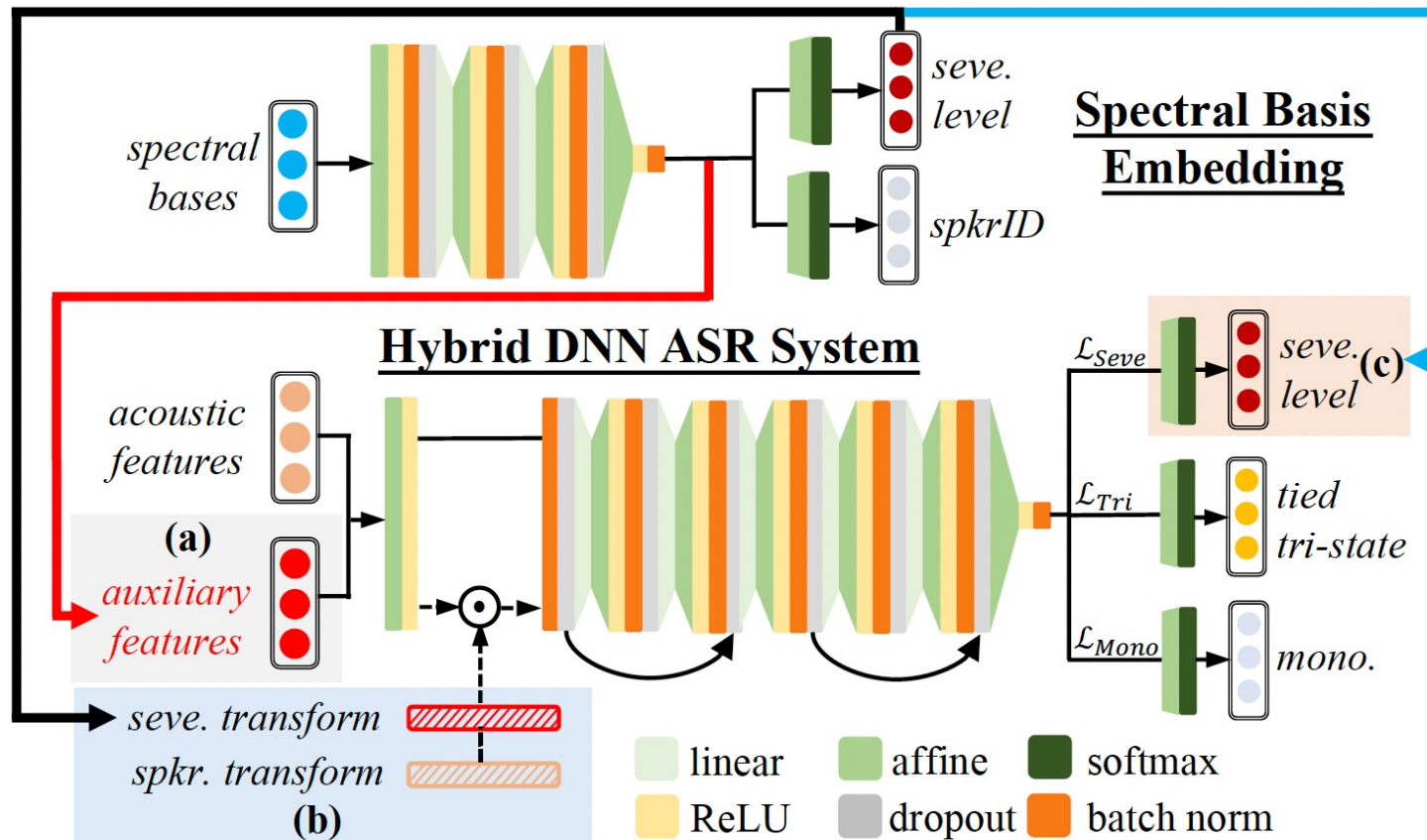
Word error rates of best published speech recognizers trained and tested using UA-Speech



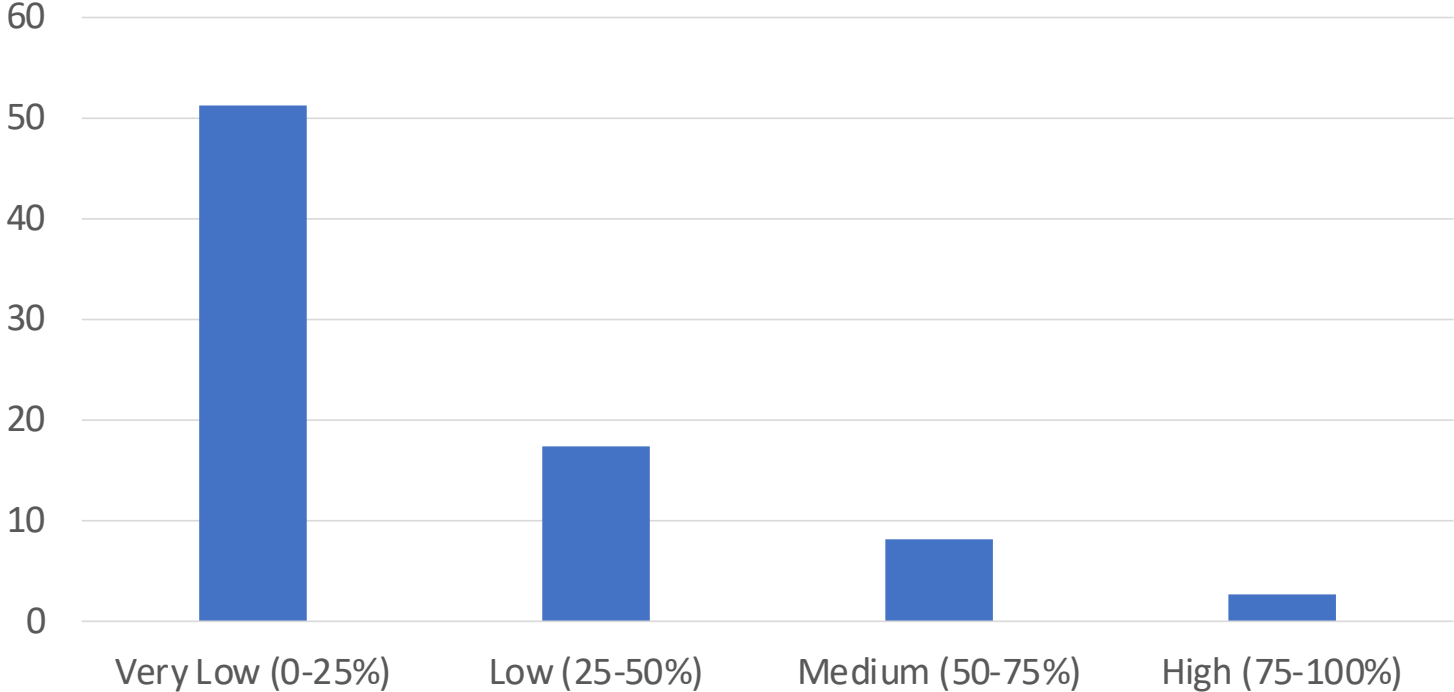
Key innovations over time

- 2014: Pool similar voices; use data from similar voices to aid speaker adaptation
- 2018: Deep neural networks
- 2019: Pitch features
- 2020: Audiovisual speech recognition during acoustic model training
- 2021: Neural network speaker adaptation
- 2022: Unsupervised pre-training of deep neural networks
- 2023: Train the neural net to also estimate speaker severity

2023: Train the neural net to also estimate dysarthria severity



Word Error Rate (%) vs. Speaker Intelligibility, UA-Speech, Best 2023 System



Intelligibility of the Speaker = Transcription Accuracy of Human Listeners

Outline

- Automatic diagnosis and analysis of voice, speech, and sleep
 - Parkinson's: Research results and remaining hurdles to clinical deployment
 - Developmental language disorder: Potential for early screening
 - Autism: Visualizations to assist clinician-parent communication
 - Infant sleep disorders: Potential for low-cost objective transcripts
 - Dysarthria intelligibility: Research results, hurdles to clinical deployment



https://commons.wikimedia.org/wiki/File:Doctor_and_patient,_1509_Wellcome_L0011744.jpg



https://commons.wikimedia.org/wiki/File:Women_practice_voting_in_Dayton_Oct._27,_1920.jpg

- Spoken language access to government, education, and employment
 - Motivation
 - Research results using the UA-Speech corpus
 - Advances since 2008 in speech technology state of the art
 - Speech Accessibility Project
 - Relevance to Speech-Language Therapy

A brief history of speech recognizer training corpora

- TIMIT (1993): 4 hours, phonetically transcribed, DARPA-sponsored
 - Broadcast News (1996): 104 hours, orthographically transcribed, DARPA-sponsored
 - Switchboard (1997): 300 hours, orthographically transcribed, DARPA-sponsored
-
- Corporate data providers enter the picture. From 2000-2020, high-cost training datasets grow in both size and quality.
-
- Librispeech (2015): 1000 hours, curated from public domain audiobooks recorded by contributors to librivox.org.
 - Multilingual Librispeech (2017): 3000 hours, curated from public domain audiobooks recorded by contributors to librivox.org.
-

Love, not money

- Audiobooks on librivox.org are usually readings of texts from Gutenberg.org.
- In order to be posted on librivox.org, the audio must be released into the public domain.
 - Portfolio samples of professional audiobook narrators
 - Books contributed by people who love them, and want them to be available
 - Languages contributed by people who love them, and want them to be available
 - Books contributed by people who love audiobooks

LibriVox
free public domain audiobooks

Search by Author, Title or Reader

Advanced search

Free public domain audiobooks

Read by volunteers from around the world.

Read

LibriVox audiobooks are read by volunteers from all over the world. Perhaps you would like to join us?

[VOLUNTEER](#)

Listen

LibriVox audiobooks are free for anyone to listen to, on their computers, iPods or other mobile device, or to burn onto a CD.

[CATALOG](#)

Welcome to Project Gutenberg

Project Gutenberg is a library of over 60,000 free eBooks

Choose among free epub and Kindle eBooks, download them or read them online. You will find the world's great literature here, with focus on older works for which U.S. copyright has expired. Thousands of volunteers digitized and diligently proofread the eBooks, for you to enjoy.

Kapellendorf
by Sophie
Hochstetter

Uncle Wiggily's AIRSHIP
by Howard R. Garis

The first church's Christmas

The old South
by Howard
Melancthon

Least Said, Soonest Mended by

X-mas Sketches from the Dartmouth
ad H

Some of our latest eBooks [Click Here for more latest books!](#)

How to use a large dataset

How much **labeled** speech does a baby hear?

- 30 (?) words/day accompanied by referential gestures = 9.1 hours of speech by age 6

How much **unlabeled** speech does a baby hear?

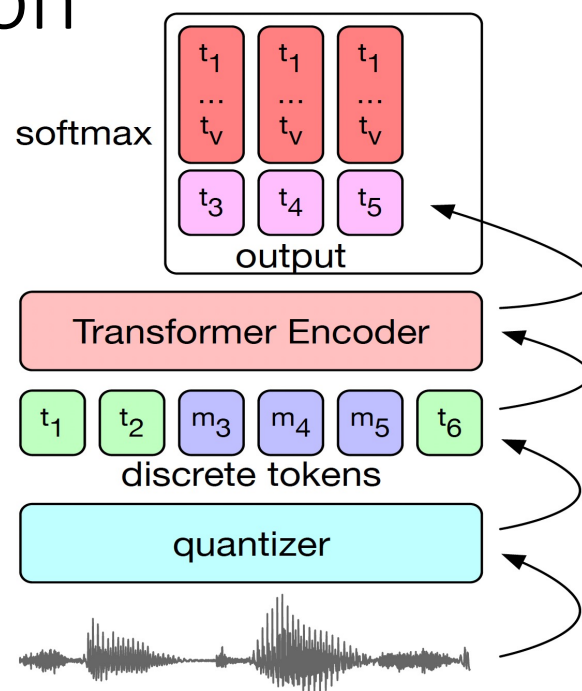
- 2000-15000 words/day = 600-4500 hours of speech by age 6 (Weisleder & Fernald, 2013)



By Steve Jurvetson from Los Altos, USA - A Proper Space Book for Babies, CC BY 2.0, <https://commons.wikimedia.org/w/index.php?curid=105132804>

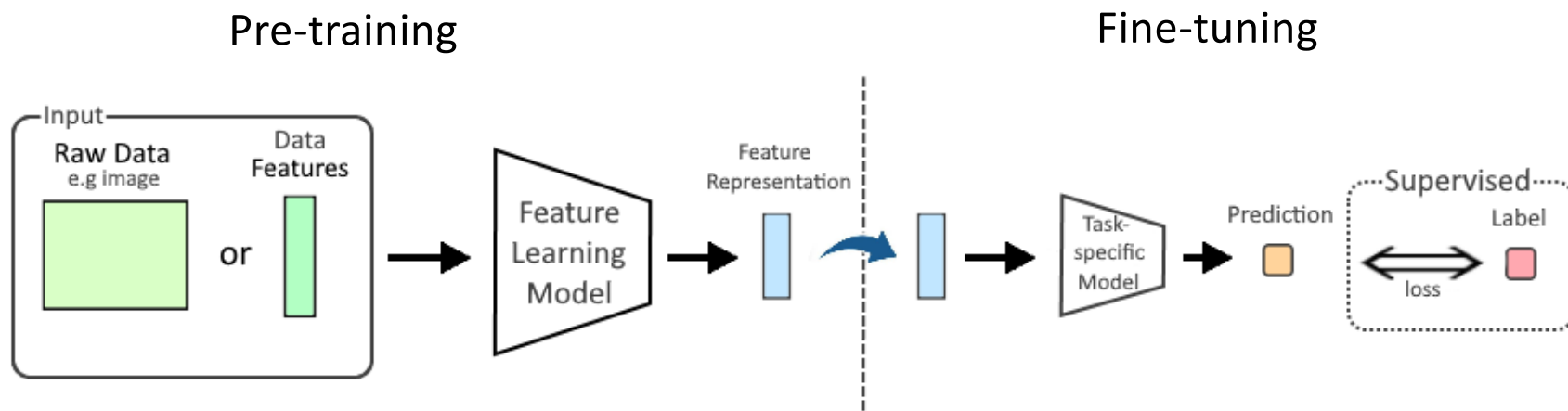
Unsupervised pre-training of transformers based on categorical perception

- Given: 60,000 hours of speech, with no associated text.
- Suppose we train the neural network to form its own categories. What would make those categories speech-like?
- **Context-Predictable Speech Categories:** given the context (the quantized units t_1 , t_2 , and t_6), it should be possible to figure out what phonemes were masked (t_3 , t_4 , t_5).



Pre-training and Fine-tuning

- A transformer is pre-trained to create its own context-predictable speech categories using, say, 60,000 hours of speech
- Then it is fine-tuned using a few hours, or a few hundred hours, or 1000 hours of labeled speech



The Speech Data Revolution: Librispeech

Panayotov et al.,

<https://doi.org/10.1109/ICASSP.2015.7178964>

- Librispeech includes 360h clean speech, 600h other speech
- Curated from librivox.org and Gutenberg.org
- Free to download, free to use and redistribute
- Librispeech is the reason for the deep-learning revolution in automatic speech recognition.

OpenSLR

Open Speech and Language Resources

[Home](#) [Resources](#)

LibriSpeech ASR corpus

Identifier: SLR12

Summary: Large-scale (1000 hours) corpus of read English speech

Category: Speech

License: CC BY 4.0

Downloads (use a mirror closer to you):

[dev-clean.tar.gz](#) [337M] (development set, "clean" speech)

Mirrors: [\[US\]](#) [\[EU\]](#) [\[CN\]](#)

[dev-other.tar.gz](#) [314M] (development set, "other", more challenging, speech) Mirrors: [\[US\]](#) [\[EU\]](#) [\[CN\]](#)

[test-clean.tar.gz](#) [346M] (test set, "clean" speech) Mirrors: [\[US\]](#) [\[EU\]](#) [\[CN\]](#)

[test-other.tar.gz](#) [328M] (test set, "other" speech) Mirrors: [\[US\]](#) [\[EU\]](#) [\[CN\]](#)

[train-clean-100.tar.gz](#) [6.3G] (training set of 100 hours "clean" speech) Mirrors: [\[US\]](#) [\[EU\]](#) [\[CN\]](#)

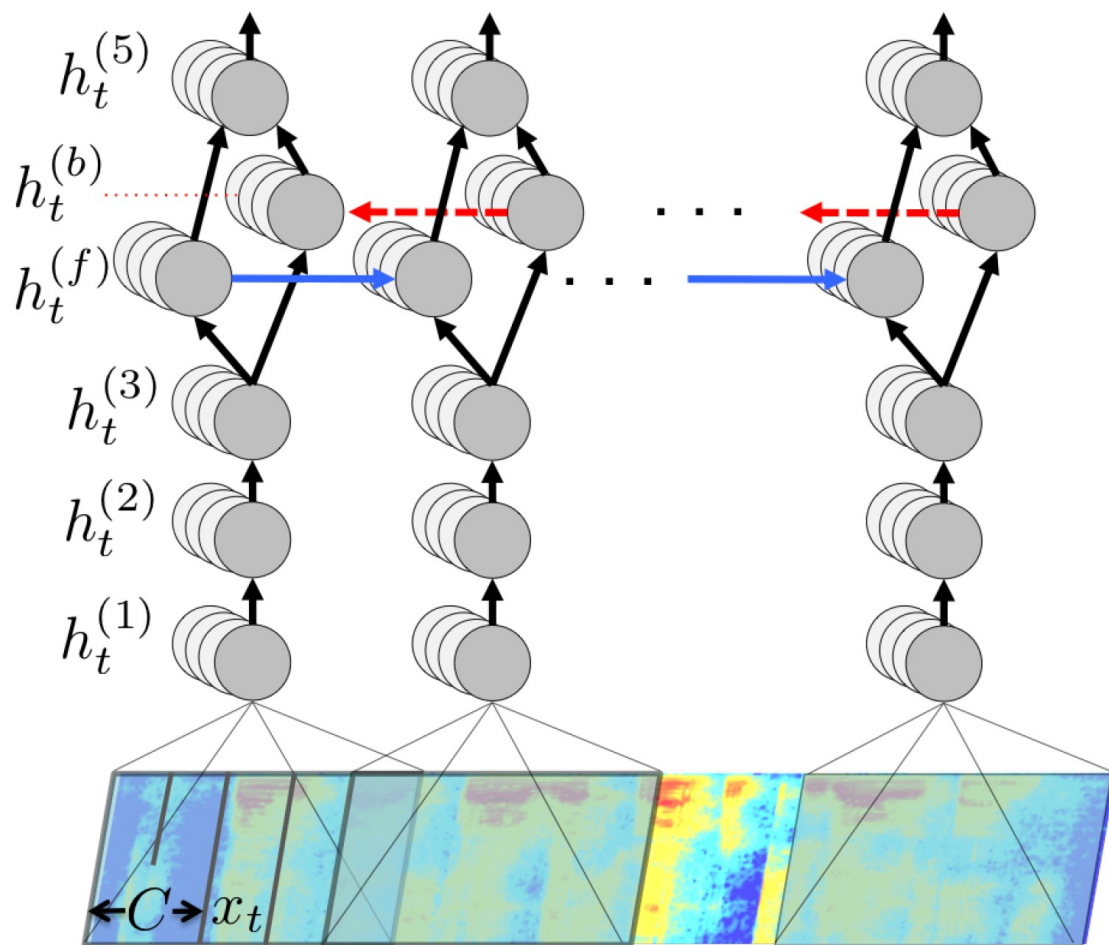
[train-clean-360.tar.gz](#) [23G] (training set of 360 hours "clean" speech) Mirrors: [\[US\]](#) [\[EU\]](#) [\[CN\]](#)

[train-other-500.tar.gz](#) [30G] (training set of 500 hours "other" speech) Mirrors: [\[US\]](#) [\[EU\]](#) [\[CN\]](#)

The resulting speech technology revolution:

End-to-end neural speech recognition

- Recurrent: academic prototypes in 2006, commercial viability in 2014
- Transformers: commercial viability in 2017
- Self-supervised pre-training: academic prototypes in 2016, commercial viability in 2020



Hannun et al., DeepSpeech, Dec 2014

Word Error Rates using Pre-Training

Pre-training makes it possible to achieve error rates of

- 4.4% using only 10 minutes of labeled data
- 2.6% using only 1 hour of labeled data

Model	Unlabeled Data	LM	dev-clean	dev-other	test-clean	test-other
<i>10-min labeled</i>						
DiscreteBERT [52]	LS-960	4-gram	15.7	24.1	16.3	25.2
wav2vec 2.0 BASE [7]	LS-960	4-gram	8.9	15.7	9.1	15.6
wav2vec 2.0 LARGE [7]	LL-60k	4-gram	6.3	9.8	6.6	10.3
wav2vec 2.0 LARGE [7]	LL-60k	Transformer	4.6	7.9	4.8	8.2
HUBERT BASE	LS-960	4-gram	9.1	15.0	9.7	15.3
HUBERT LARGE	LL-60k	4-gram	6.1	9.4	6.6	10.1
HUBERT LARGE	LL-60k	Transformer	4.3	7.0	4.7	7.6
HUBERT X-LARGE	LL-60k	Transformer	4.4	6.1	4.6	6.8
<i>1-hour labeled</i>						
DeCoAR 2.0 [51]	LS-960	4-gram	-	-	13.8	29.1
DiscreteBERT [52]	LS-960	4-gram	8.5	16.4	9.0	17.6
wav2vec 2.0 BASE [7]	LS-960	4-gram	5.0	10.8	5.5	11.3
wav2vec 2.0 LARGE [7]	LL-60k	Transformer	2.9	5.4	2.9	5.8
HUBERT BASE	LS-960	4-gram	5.6	10.9	6.1	11.3
HUBERT LARGE	LL-60k	Transformer	2.6	4.9	2.9	5.4
HUBERT X-LARGE	LL-60k	Transformer	2.6	4.2	2.8	4.8

Outline

- Automatic diagnosis and analysis of voice, speech, and sleep
 - Parkinson's: Research results and remaining hurdles to clinical deployment
 - Developmental language disorder: Potential for early screening
 - Autism: Visualizations to assist clinician-parent communication
 - Infant sleep disorders: Potential for low-cost objective transcripts
 - Dysarthria intelligibility: Research results, hurdles to clinical deployment



https://commons.wikimedia.org/wiki/File:Doctor_and_patient,_1509_Wellcome_L0011744.jpg



https://commons.wikimedia.org/wiki/File:Women_practice_voting_in_Dayton_Oct._27,_1920.jpg

- Spoken language access to government, education, and employment
 - Motivation
 - Research results using the UA-Speech corpus
 - Advances since 2008 in speech technology state of the art
 - Speech Accessibility Project
 - Relevance to Speech-Language Therapy

The Speech Accessibility Project

- Speech technology is now good enough to be useful for people speaking General American English and Received Pronunciation without disabilities.
- To make it useful for people with disabilities, we need about 1000 hours of transcribed speech (~1.2 million sentences).



SPEECH ACCESSIBILITY PROJECT

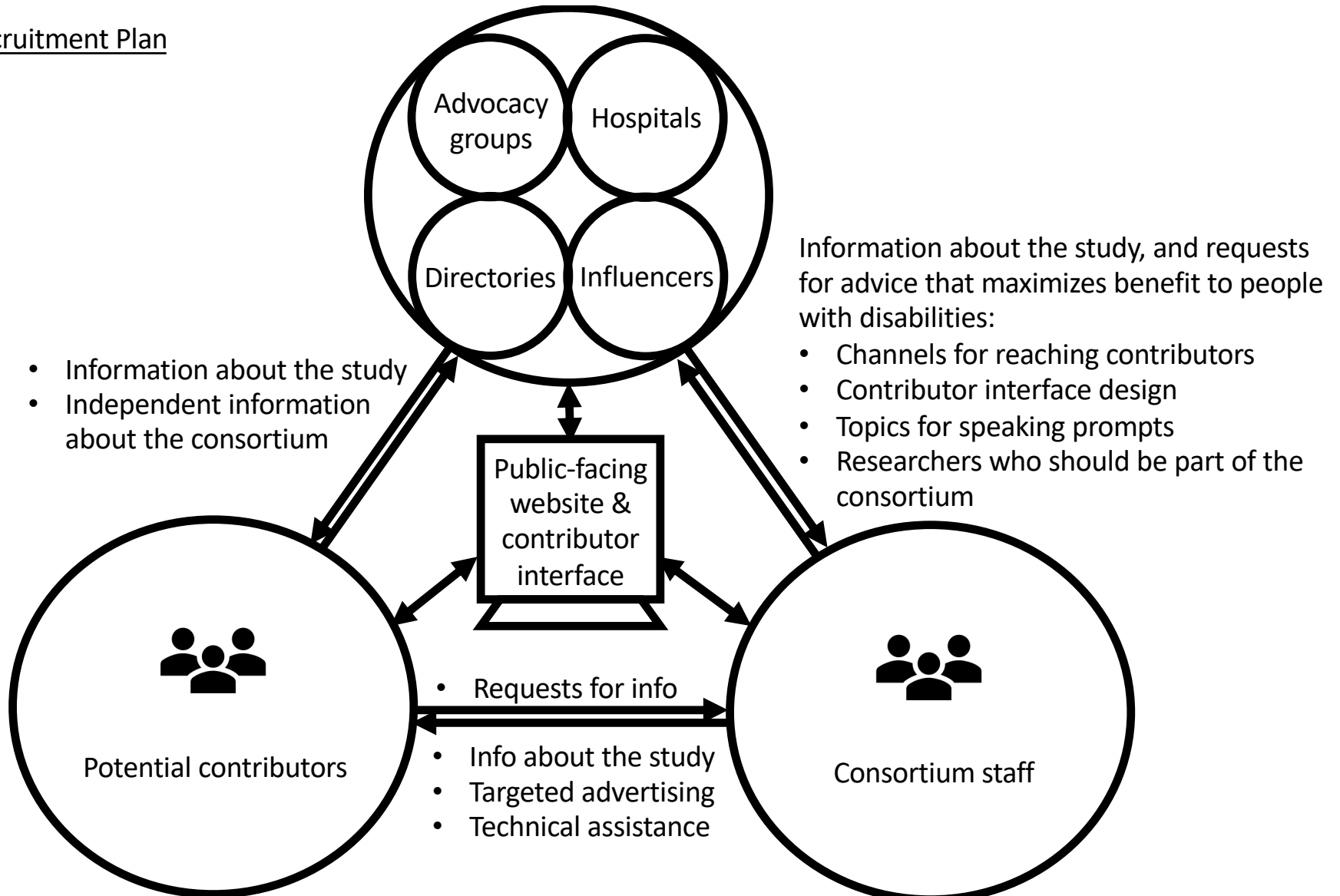
Beckman Institute for Advanced Science and Technology

The Speech Accessibility Project is intended as a communication channel that connects people with disabilities with the engineers and computer scientists who create speech technology.

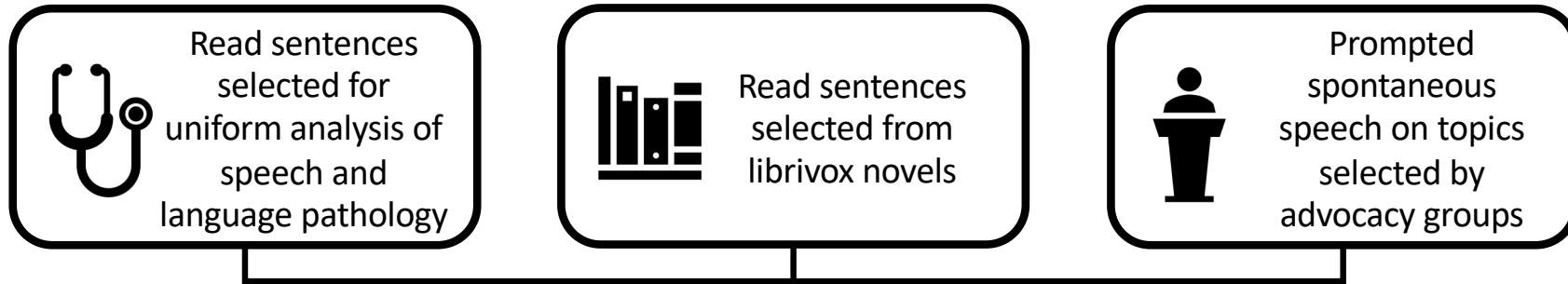
Recruitment strategy and status

- Target: 1000 hours of transcribed speech
 - 2000 participants, 600 sentences per person
 - Recruiting only in the United States
- Recruitment strategy: 5 etiologies, 400 people each
 - Parkinson's, ALS, Down Syndrome, Stroke, Cerebral Palsy
 - Other etiologies will be screened if they volunteer
 - Currently recruiting: Etiology 1, Parkinson's
- Consent process and mentoring
 - Potential participants (746 as of 2023/10/10) meet a speech-language therapist online, who decides whether their speech is sufficiently affected to meet the needs of the project (283 as of 2023/10/10)

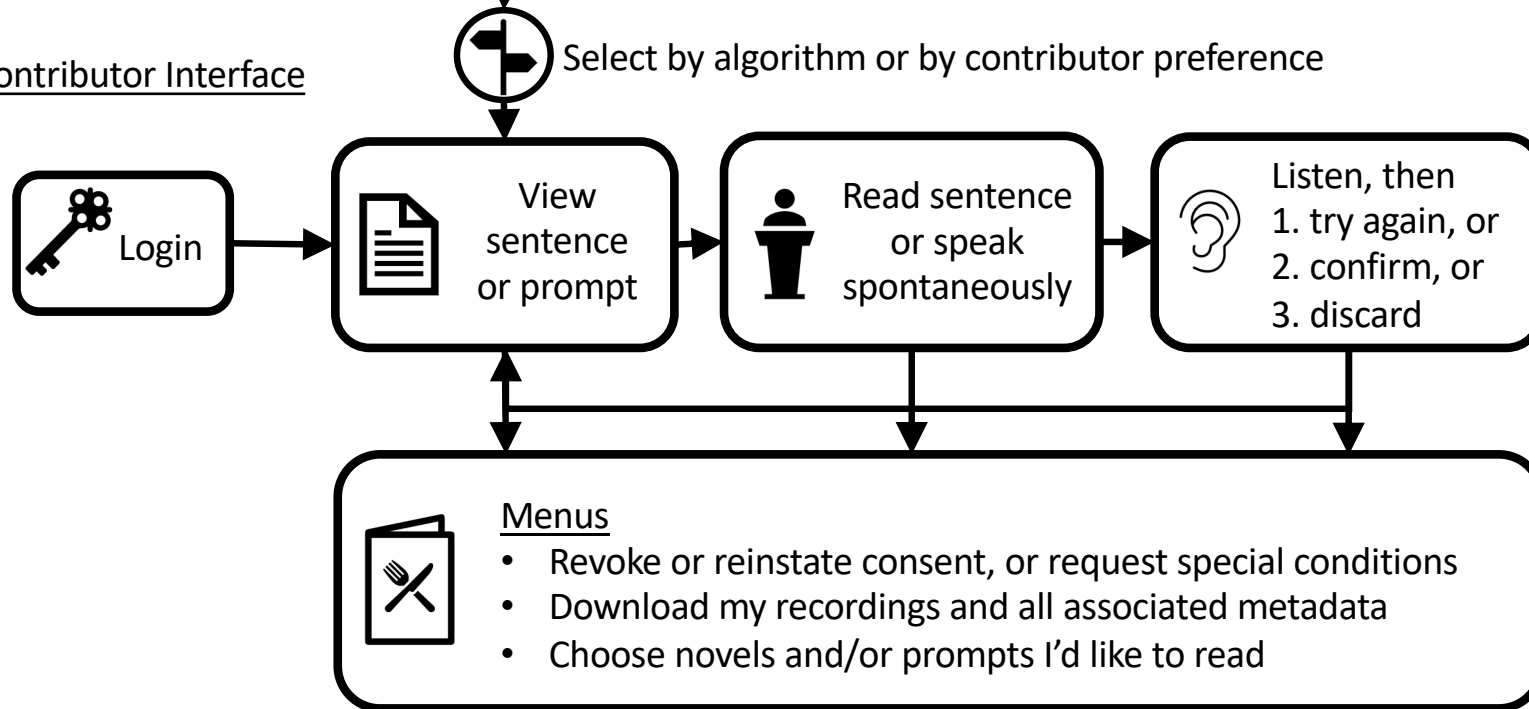
Outreach & Recruitment Plan



Prompt Sources



Contributor Interface



Distribution to researchers

- The data distribution includes:
 - Contributed audio
 - Prompt text, Corrected text,
 - Perceptual dysarthria attributes (etiology-dependent subset of Darley, Aronson & Brown perceptual scales, 6 recordings/participant annotated by one speech pathologist)
- Data is packaged in zip files on secure cloud storage
- Members of the original consortium (Amazon, Apple, Google, Meta, Microsoft) have already started testing ASR using subsets of data from people with Parkinson's
- Data will be released to other research organizations (companies, universities, and other organizations) 6-12 months after release to consortium members

Human subject protection principles

Contributors should know how their data will be used;
Researchers should commit to the same terms.

- The purpose of the study and permitted uses, as specified in the participant consent form (<https://github.com/speechaccessibility/FAQ>),
- ...are the same as the Permitted Purpose and permitted uses in the data use agreement.



University of Illinois Urbana-Champaign
Online Consent Form



IRB Number: 23183
IRB Approval Date: 03/07/2023
IRB Expiration Date: 03/02/2028

Speech Accessibility Project: Individuals with disabilities helping researchers to improve technology

You are being asked to participate in a voluntary research study that is called the “Speech Accessibility Project.” The purpose of this study is to help researchers at universities and companies to develop spoken-language user interfaces that work for people with atypical speech. Software programs that understand speech are developed using machine learning. Machine learning is a software development method that imitates the way human behavior whenever it makes a mistake in the future. By providing examples of possible for the software to learn how to behave in the future. Participating in this study involves providing your audio recordings to a secure server for analysis. This process takes about 5 samples or about 5 hours of your time. The expected benefit of this research to you and organizations using your speech is to improve communication with speech and motor disabilities, w

Project Name: The Speech Accessibility Project
<https://speechaccessibilityproject.be>

LICENSED SPEECH SIGNAL DATA DATA USE TERMS AND CONDITIONS

The licenses to Licensed Speech Signal Data granted to each Member pursuant to Section 9.b. of the Agreement are subject to the following Terms and Conditions. Each Member receiving such a license is referred to herein as the “Data User”. Any capitalized terms used but not otherwise defined in this Exhibit C shall have the meanings given to them elsewhere in the Agreement.

1. Provision and Use of Data

1.1 Effective as of the Effective Date, Data User may make, have made, use, load, access, store, copy, reproduce, destroy, modify, transmit, display, make derivative works of and otherwise use the Data made available to it by Organization under the Agreement on a non-exclusive, non-assignable, non-sublicensable (except as provided in Section 2.1 below) basis solely for the Permitted Purpose subject to the terms of this Exhibit C and the Agreement. Any such modifications or derivative works made by Data User are “**Modification(s)**”.

1.2 This Agreement does not restrict Data User’s use or modification of any portions of the Data that Organization makes publicly available under a more permissive license, if applicable, to the extent Data User uses or modifies the Data under the terms of such public license, or that otherwise become public.

1.3 The rights and restrictions set forth in this Exhibit C as it relates to the use, modification, distribution, disclosure, and privacy of the Data may only be modified by a written agreement between Data User and Organization specifically indicating it is amending these Terms and Conditions, e.g., these Terms and Conditions may not be amended by or superseded by any general terms and conditions that are presented or included as part of access to or use of data storage, data processing, platform, or computing resources.

How we protect participant privacy

- Every recording verified by a human being before distribution
 - Personally identifiable information zeroed out if necessary
- Researchers sign data use agreement
- Researchers get no participant-provided data or metadata except the speech itself
 - UIUC team supplies prompts, text transcript & speech pathology annotations



University of Illinois Urbana-Champaign
Online Consent Form



IRB Number: 23183
IRB Approval Date: 03/07/2023
IRB Expiration Date: 03/02/2028

Speech Accessibility Project: Individuals with disabilities helping researchers to improve technology

You are being asked to participate in a voluntary research study that is called the “Speech Accessibility Project.” The purpose of this study is to help researchers at universities and companies to develop spoken-language user interfaces that work for people with atypical speech. Software programs that understand speech are developed using machine learning. Machine learning is a software development method that imitates the way human behavior whenever it makes a mistake in the future. By providing examples of human behavior, it is possible for the software to learn how to avoid making the same mistakes in the future. Participating in this study will require you to provide your audio recordings to a secure server for analysis. This will take about 5 samples or about 5 hours of your time and is very convenient. Risks associated with this research are minimal. The expected benefit of this research to you and to other individuals with speech and motor disabilities, will be significant.

Project Name: The Speech Accessibility Project
<https://speechaccessibilityproject.be>

LICENSED SPEECH SIGNAL DATA DATA USE TERMS AND CONDITIONS

The licenses to Licensed Speech Signal Data granted to each Member pursuant to Section 9.b. of the Agreement are subject to the following Terms and Conditions. Each Member receiving such a license is referred to herein as the “Data User”. Any capitalized terms used but not otherwise defined in this Exhibit C shall have the meanings given to them elsewhere in the Agreement.

1. Provision and Use of Data

1.1 Effective as of the Effective Date, Data User may make, have made, use, load, access, store, copy, reproduce, destroy, modify, transmit, display, make derivative works of and otherwise use the Data made available to it by Organization under the Agreement on a non-exclusive, non-assignable, non-sublicensable (except as provided in Section 2.1 below) basis solely for the Permitted Purpose subject to the terms of this Exhibit C and the Agreement. Any such modifications or derivative works made by Data User are “**Modification(s)**”.

1.2 This Agreement does not restrict Data User’s use or modification of any portions of the Data that Organization makes publicly available under a more permissive license, if applicable, to the extent Data User uses or modifies the Data under the terms of such public license, or that otherwise become public.

1.3 The rights and restrictions set forth in this Exhibit C as it relates to the use, modification, distribution, disclosure, and privacy of the Data may only be modified by a written agreement between Data User and Organization specifically indicating it is amending these Terms and Conditions, e.g., these Terms and Conditions may not be amended by or superseded by any general terms and conditions that are presented or included as part of access to or use of data storage, data processing, platform, or computing resources.

DUA key provisions

- Researchers may not identify participants
- Researchers may not redistribute data outside of the researcher's organization
 - Except that short samples may be played in demos online and conferences if the participants checks an optional agreement checkbox
- Researchers must have organizational, physical, and software protections against data theft



University of Illinois Urbana-Champaign
Online Consent Form



IRB Number: 23183
IRB Approval Date: 03/07/2023
IRB Expiration Date: 03/02/2028

Speech Accessibility Project: Individuals with disabilities helping researchers to improve technology

You are being asked to participate in a voluntary research study that is called the "Speech Accessibility Project." The purpose of this study is to help researchers at universities and companies to develop spoken-language user interfaces that work for people with atypical speech. Software programs that understand speech are developed using machine learning. Machine learning is a software development method that imitates the way human behavior whenever it makes a mistake in the future. By providing examples of possible for the software to learn how to behave in the future. Participating in this study will require you to provide audio recordings to a secure server for about 5 hours of your time and convenience. Risks associated with this research to you and organizations using your speech with speech and motor disabilities, w

Project Name: The Speech Accessibility Project
<https://speechaccessibilityproject.be>

LICENSED SPEECH SIGNAL DATA DATA USE TERMS AND CONDITIONS

The licenses to Licensed Speech Signal Data granted to each Member pursuant to Section 9.b. of the Agreement are subject to the following Terms and Conditions. Each Member receiving such a license is referred to herein as the "Data User". Any capitalized terms used but not otherwise defined in this Exhibit C shall have the meanings given to them elsewhere in the Agreement.

1. Provision and Use of Data

1.1 Effective as of the Effective Date, Data User may make, have made, use, load, access, store, copy, reproduce, destroy, modify, transmit, display, make derivative works of and otherwise use the Data made available to it by Organization under the Agreement on a non-exclusive, non-assignable, non-sublicensable (except as provided in Section 2.1 below) basis solely for the Permitted Purpose subject to the terms of this Exhibit C and the Agreement. Any such modifications or derivative works made by Data User are "**Modification(s)**".

1.2 This Agreement does not restrict Data User's use or modification of any portions of the Data that Organization makes publicly available under a more permissive license, if applicable, to the extent Data User uses or modifies the Data under the terms of such public license, or that otherwise become public.

1.3 The rights and restrictions set forth in this Exhibit C as it relates to the use, modification, distribution, disclosure, and privacy of the Data may only be modified by a written agreement between Data User and Organization specifically indicating it is amending these Terms and Conditions, e.g., these Terms and Conditions may not be amended by or superseded by any general terms and conditions that are presented or included as part of access to or use of data storage, data processing, platform, or computing resources.

https://github.com/speechaccessibility/AudioControls

Search or jump to... Pull requests Issues Codespaces Marketplace Explore

speechaccessibility / AudioControls Public Edit Pins Unwatch 1

Code Issues Pull requests Actions Projects Wiki Security Insights S

main 1 branch 7 tags Go to file Add file Code

Commit	Author	Message	Time
792ad73	skremitzki	dispatch error events in catch blocks	yesterday
			13 commits
AudioControls.ts		dispatch error events in catch blocks	yesterday
LICENSE		Added more specific copyright holder information.	5 months ago
README.md		modified Readme to remove the references to the various "Play...	2 weeks ago

README.md

AudioControls

Typescript implementation of audio recording, playback and simulated waveform display based on the MediaRecorder API.

Example:

```
let recordButton = document.getElementById('recordButton')
recordButton.addEventListener(
  'AudioControls.RecordingStarted',
```

<https://github.com/speechaccessibility/AudioControls>

- Software design principles:
 - Contributor safety
 - Data recoverability
 - Contributor ease of use
 - Annotator ease of use
- Contributor and Annotator user interfaces go through dev and test environments before being pushed to production.
- All finished products are released open-source.

Outline

- Automatic diagnosis and analysis of voice, speech, and sleep
 - Parkinson's: Research results and remaining hurdles to clinical deployment
 - Developmental language disorder: Potential for early screening
 - Autism: Visualizations to assist clinician-parent communication
 - Infant sleep disorders: Potential for low-cost objective transcripts
 - Dysarthria intelligibility: Research results, hurdles to clinical deployment



https://commons.wikimedia.org/wiki/File:Doctor_and_patient,_1509_Wellcome_L0011744.jpg



https://commons.wikimedia.org/wiki/File:Women_practice_voting_in_Dayton_Oct._27,_1920.jpg

- Spoken language access to government, education, and employment
 - Motivation
 - Research results using the UA-Speech corpus
 - Advances since 2008 in speech technology state of the art
 - Speech Accessibility Project
 - Relevance to Speech-Language Therapy

Relevance to Speech-Language Therapy

- Your patients get improved access to democracy, education & employment
- Speech-controlled environmental control systems (Hawley et al., <https://doi.org/10.1016/j.medengphy.2006.06.009>)
- Use of an automatic speech recognizer is a type of speech practice
- Automatic objective pronunciation testing uses ASR (Witt & Young, [https://doi.org/10.1016/S0167-6393\(99\)00044-8](https://doi.org/10.1016/S0167-6393(99)00044-8))

Summary of Key Results

Automatic diagnosis and analysis of voice, speech, and sleep:

- Parkinson's: Automatic diagnosis is possible with 79-94% accuracy, highly variable depending on the testing dataset; larger datasets are needed
- Developmental language disorder: Rapid screening tests might count the number of distinct subject-verb combinations and use the result to recommend formal diagnosis by an SLT
- Autism: Visualizations of child's gaze, speech and gesture can assist clinician-parent communication
- Infant sleep disorders: Wearable devices can list waking incidents, may soon list REM vs non-REM sleep
- Dysarthria intelligibility: There are many studies, using few datasets. The standard data split, developed for ASR, has the same participants in training & test corpora; there is growing awareness of the need for distinct train/test participants, but not enough data.

Summary of Key Results

Spoken language access to government, education, and employment:

- Speech technology has the potential to give people with disabilities better access to government, education and employment.
- Informal international research competition using the UA-Speech corpus has resulted in word error rates that drop by 9.4% (relative) per year, for the past 15 years running.
- Advances since 2008 in speech technology state of the art have been built on top of 1000 hours of transcribed speech (librispeech) and many thousands of hours of untranscribed speech, made available to the world by people who love books.
- The Speech Accessibility Project seeks to build a bridge between the disability community and the technology community, for the purpose of facilitating similar advances in speech accessibility. For now, that bridge takes the form of payment to participants in exchange for speech recordings, which we curate and distribute to researchers who promise to protect participant privacy.