University at Buffalo *The State University of New York* | REACHING OTHERS

Progress toward estimating the minimal clinically important difference of intelligibility: A crowdsourced perceptual experiment

Kaila L. Stipancic[1], Frits van Brenk[1,2], Mengyang Qiu[1], & Kris Tjaden[1]

[1]Motor Speech Disorders Laboratory, Department of Communicative Disorders and Sciences, University at Buffalo; [2]Utrecht Institute of Linguistics OTS, Utrecht University

klstip@buffalo.edu

INTRODUCTION

- The **minimal clinically important difference (MCID)** has been used widely across the rehabilitation sciences field to define the smallest amount of change in a particular outcome measure that patients (or others) perceive as beneficial [1]
- The MCID is a necessary supplement to the **minimally detectable change (MDC)** which is the smallest amount of change outside of measurement error, which is necessary for us to be confident that a change is real [2]
  - The MDC has been previously established for sentence intelligibility in healthy control speakers and in speakers with ALS [3], MS, and PD [4]
  - For speakers with mild dysarthria, the MDC of intelligibility is between 3-6% [3,4]
  - BUT the MDC does not indicate what is *clinically* meaningful

The **purpose** of this study was to estimate the minimal clinically important difference (MCID) of speech intelligibility in neurologically healthy controls and speakers with dysarthria due to multiple sclerosis (MS) and Parkinson's disease (PD) as determined by non-expert listeners in the presence of background noise.

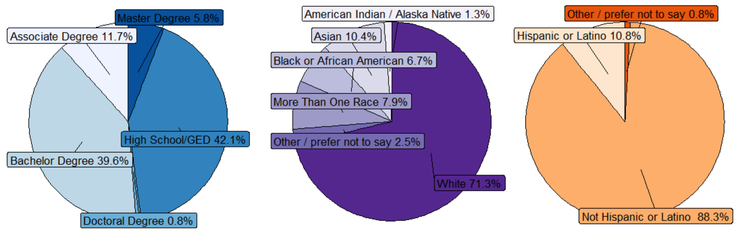
METHODS

Speakers

- 16 neurologically healthy controls, 16 speakers with MS, and 16 speakers with PD from [5, 6] = 48 speakers total
- Audio recorded reading aloud a subset of Harvard psychoacoustic sentences [7] in habitual, clear, fast, loud, and slow speaking conditions
- Audio signals were normalized for peak amplitude and mixed with multitalker babble to reduce ceiling effects and enhance intelligibility differences between conditions, which is desirable for estimating the MCID
- 48 speakers x 10 condition-combinations (i.e., habitual-clear, habitual-fast, clear-fast, etc.) = 480 samples of 6 sentences (3 in each condition) – divided into 24 lists with 20 samples each
  - Repeated x2 condition-combinations for intra-rater reliability

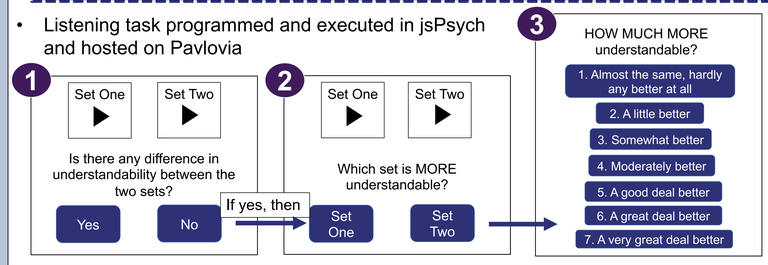
Listeners

- 240 crowdsourced listeners participated in this study via Prolific (prolific.co)
  - Average age: 24.13 years (SD = 3.66, range = 18-30)
  - 170 female, 55 male, 9 other/prefer not to say, 5 unspecified, 1 unknown



Procedures

- Listening task programmed and executed in jsPsych and hosted on Pavlovia



- Each listener heard 1 list of 20 samples
  - 10 listeners/list
- Transcription intelligibility previously obtained in [5]
- Anchor scale for calculating the MCID = global ratings of change scale (GROC per [1]):
  - No change = 0
  - Small changes = 1-3
  - Moderate changes = 4-5
  - Large changes = 6-7

Data analysis

- Reliability calculated for each of the 3 questions in the listening task (both intra- and inter-rater) using Fleiss' Kappa and Intraclass Correlation Coefficients (ICCs)
- Following methods in [3], receiving operating characteristic (ROC) curves used to determine how well the change in intelligibility scores between speaking conditions differentiated between those speakers for whom listeners identified a change in understandability and those for whom listeners did not identify a change in understandability
  - MCID typically defined as the ROC threshold that maximizes both sensitivity and specificity [1]
  - Area under the curve (AUC) calculated to establish the probability of intelligibility scores to distinguish between perceptibly changed and unchanged speakers
- Calculated the average intelligibility percentage difference between conditions for each point on the GROC scale and identified the sensitivity, specificity, and accuracy of that percentage for distinguishing between changed and unchanged speakers
  - Conducted ANOVAs between levels of the GROC

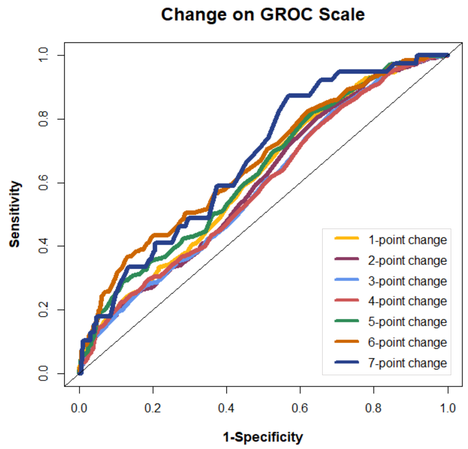
RESULTS

3 levels of reliability

Question	Type of reliability	Reliability statistic	Result
1 Is there a difference in understandability? Yes vs. No	Intra-rater	Fleiss' Kappa	.27
	Inter-rater	Fleiss' Kappa	.14
2 Which stimuli are more understandable? One vs. Two	Intra-rater	Fleiss' Kappa	.46
	Inter-rater	Fleiss' Kappa	.36
3 How much more understandable? 7-point scale	Intra-rater	ICC3k	.55
	Inter-rater	ICC3k	.75

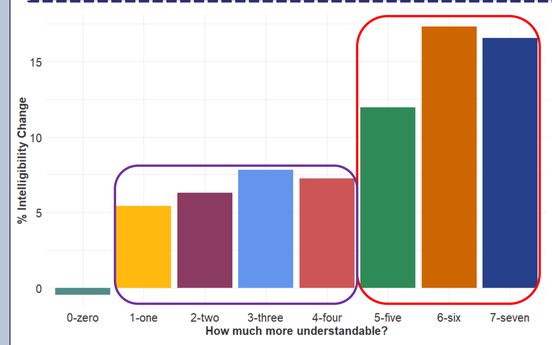
all sig at $p < .001$, except #1 inter-rater reliability: 4 lists sig at $p < .05$ and 4 lists n.s.

ROC curves



Change on GROC	AUC (95% CI)	Threshold
1	.61 (.60-.63)	-3.17
2	.60 (.58-.61)	-3.17
3	.59 (.57-.60)	-3.17
4	.59 (.57-.61)	-3.17
5	.63 (.51-.66)	-1.33
6	.66 (.62-.70)	26.67
7	.67 (.59-.75)	0.67

Average intelligibility change



All comparisons significantly different at $p < .001$ except:

- 1-2; $p = .979$
- 1-3; $p = .222$
- 1-4; $p = .672$
- 2-3; $p = .655$
- 2-4; $p = .975$
- 3-4; $p = .999$
- 3-7; $p = .054$
- 5-6; $p = .058$
- 5-7; $p = .786$
- 6-7 ; $p = .999$

Change on GROC	Mean Intelligibility	SD	Closest threshold (ROCs)	Specificity	Sensitivity	Accuracy	
0	-0.46	18.13	N/A	N/A	N/A	N/A	"no change" = 0%
1	5.40	17.24	5.33	.66	.43	.50	"small change" = 8.5%
2	6.28	17.47	6.33	.65	.42	.52	
3	7.81	17.01	8.00	.67	.39	.57	
4	7.25	16.96	7.33	.66	.40	.61	
5	11.95	18.30	11.33	.74	.39	.71	"moderate-large change" = 15%
6	17.32	21.08	17.33	.80	.43	.79	
7	16.58	19.79	16.67	.79	.41	.79	

SUMMARY AND CONCLUSIONS

- MDC of intelligibility previously calculated for mildly impaired speakers with MS and PD in [4] = between 3-6%; MCID must be larger in magnitude to be meaningful
  - MCID of intelligibility calculated here:
    - **Small change = 8.5%**
    - **Moderate-large change = 15%**
- Future directions: MCID should be calculated for each context in which intelligibility is used as an outcome measure (i.e., across patient populations, types of listeners, methods used, etc.)

Preliminary conclusions:

1. Demonstrates feasibility of the novel experimental paradigm for collecting crowdsourced perceptual data for estimating MCIDs.
2. Provides empirical evidence that clinical tools for the perception of intelligibility by everyday listeners should have only 3 categories ("no change", "a little bit of change", "a moderate/large amount of change").
3. A critical step toward development of a universal language with which to evaluate changes in intelligibility as a result of speech-language therapy and disease progression.

REFERENCES

1. Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials, 10*(4), 407-415.; 2. Riddle, D., & Stratford, P. (2013). *Is this change real? Interpreting patient outcomes in physical therapy.* Philadelphia, PA: Davis Plus.; 3. Stipancic, K. L., Yunusova, Y., Berry, J. D., & Green, J. R. (2018). Minimally detectable change and minimal clinically important difference of a decline in sentence intelligibility and speaking rate for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research, 61*(11), 2757–2771.; 4. Stipancic, K. L., & Tjaden, K. (2022). Minimally detectable change of intelligibility in speakers with multiple sclerosis and Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 65*, 1858-1866.; 5. Sussman, J. E., & Tjaden, K. (2012). Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research, 55*, 1208-1219.; 6. Tjaden, K., Sussman, J. E., & Wilding, G. E. (2014). Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research, 57*, 779-792.; 7. The Institute of Electrical and Electronics Engineers. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17*, 225–246.

Research supported by NIH R01 DC004689 (PI: Tjaden).
No relevant non-financial relationships to disclose.