



PEGASAS Project 35

Enhanced Hands-Minimized Weather Interfaces for Pilots

Barrett Caldwell, PhD

Mary E. Johnson, PhD

Brandon J. Pitts, PhD

Shantanu Gupta

Purdue University

Deborah S. Carstens, PhD

Michael Splitt, MS

Faraz Amiraslani

Joshua Breininger

Rusheel Raj Panakadan

Florida Institute of Technology

Project Update, June 2024

Although the FAA has sponsored this project, it neither endorses nor rejects the findings of this research. The presentation of this information is in the interest of invoking technical community comment on the results and the conclusions of the research.

Deliverable AWIP.X.1: Metric(s) to quantify cadence of PIREP submissions as affected by experience, stress, and environmental conditions

- ✓ Literature Review to finalize four metrics to quantify speech cadence
 - ✓ Words Per Minute (WPM)
 - ✓ Syllables Per Minute (SPM)
 - ✓ Phonation-Time Ratio
 - ✓ Number of pauses per duration
- ✓ Compute cadence metrics for three PIREP datasets
 - ✓ BP Goal Data – 18 PIREPs
 - ✓ FIT Data – 258 PIREPs (Short, Average, Long)
 - ✓ Purdue Data – 120 PIREPs (Pilot, Non-Pilot)
- ✓ Evaluate cadence metrics against Word Error Rate (WER) of PIREP tool
 - ✓ Exploratory analysis
 - ✓ Regression modelling
 - ✓ Other data science approaches

Deliverable AWIP.X.2: Comparison between remotely submitted (from AWIP.3.2), pre-existing (previously recorded from FAA or other sources), and in-person PIREP datasets to verify ML performance.

- ✓ Set up independent PIREP Submittal Tool at FIT
 - ✓ Transfer files and instructions from Purdue to FIT
 - ✓ Install required software and tool
 - ✓ Troubleshoot errors and version control
 - ✓ Independent copy of tool running at FIT since 04/05
- ✓ Design experiment to compare performance of PIREP Submittal Tool
 - ✓ Run selected files at Purdue and FIT
 - ✓ String comparison of PIREP transcription
 - ✓ String comparison of PIREP codes

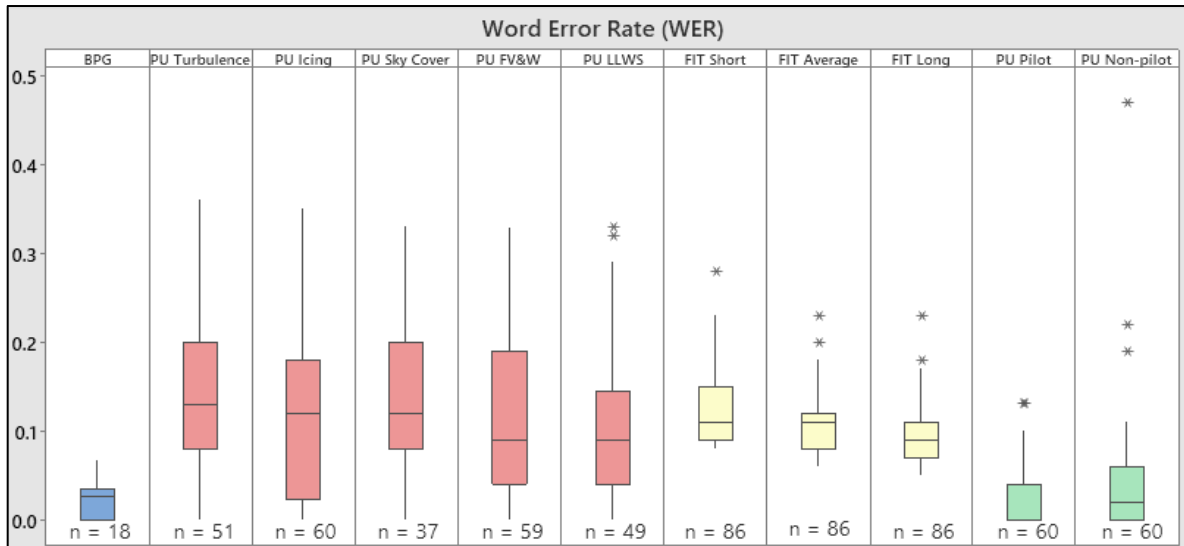
Deliverable AWIP.X.3: Report of experimental results and completion of MinWxSvc recommendation

Selected Error Metric: Word Error Rate

Word Error Rate (WER)

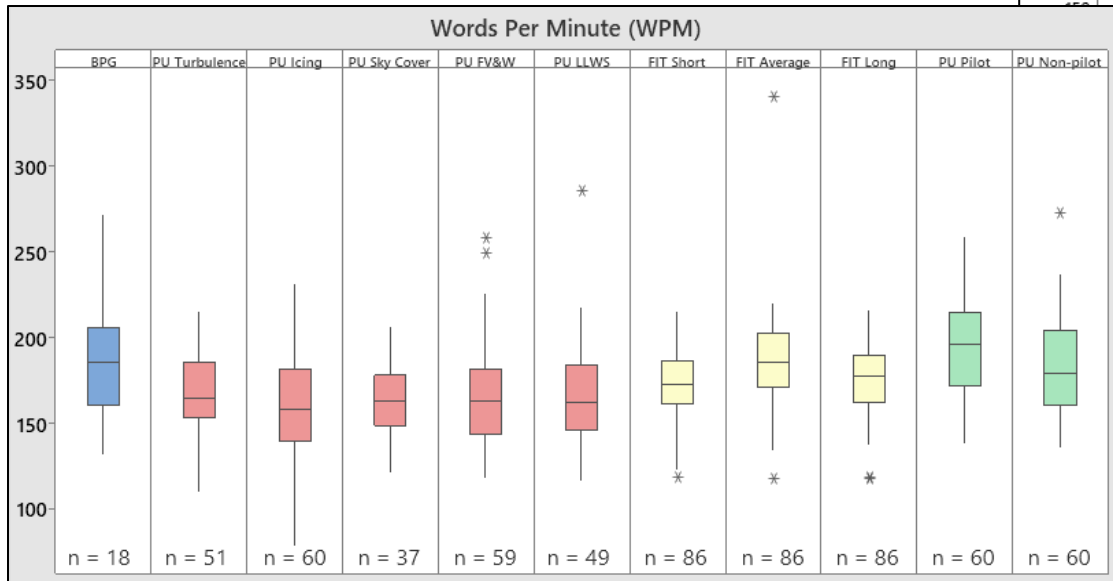
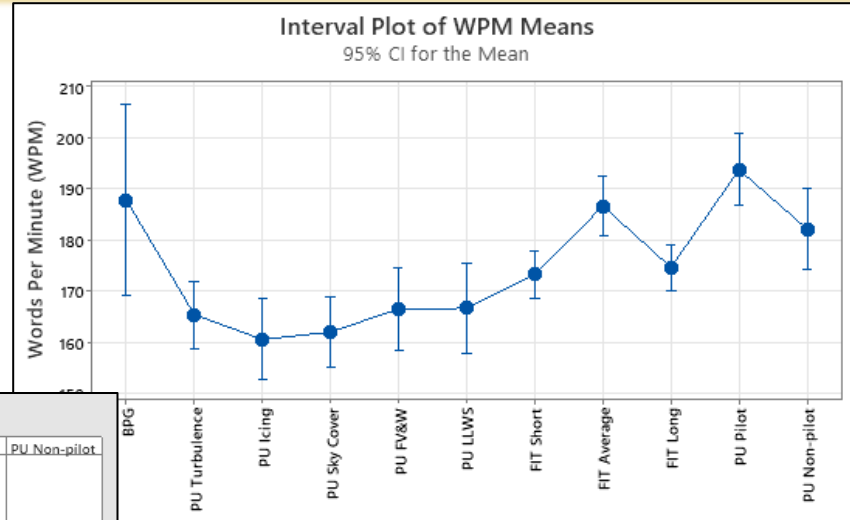
- Error rate in the speech-to-text portion of the PIREP submittal tool
- $$\frac{\text{Substitutions (S)} + \text{Deletions (D)} + \text{Insertions (I)}}{N1 = \text{Hits (H)} + \text{Substitutions (S)} + \text{Deletions (D)}}$$

Datasets Evaluated	
1. BPG: Business Plan Goal	
2. Purdue 282 PIREPs	1. PU Turbulence 2. PU Icing 3. PU Sky Cover 4. PU FV&W
3. FIT 258 PIREPs	1. FIT Short 2. FIT Average 3. FIT Long
4. Purdue 120 PIREPs	1. PU Pilot 2. PU Non-Pilot



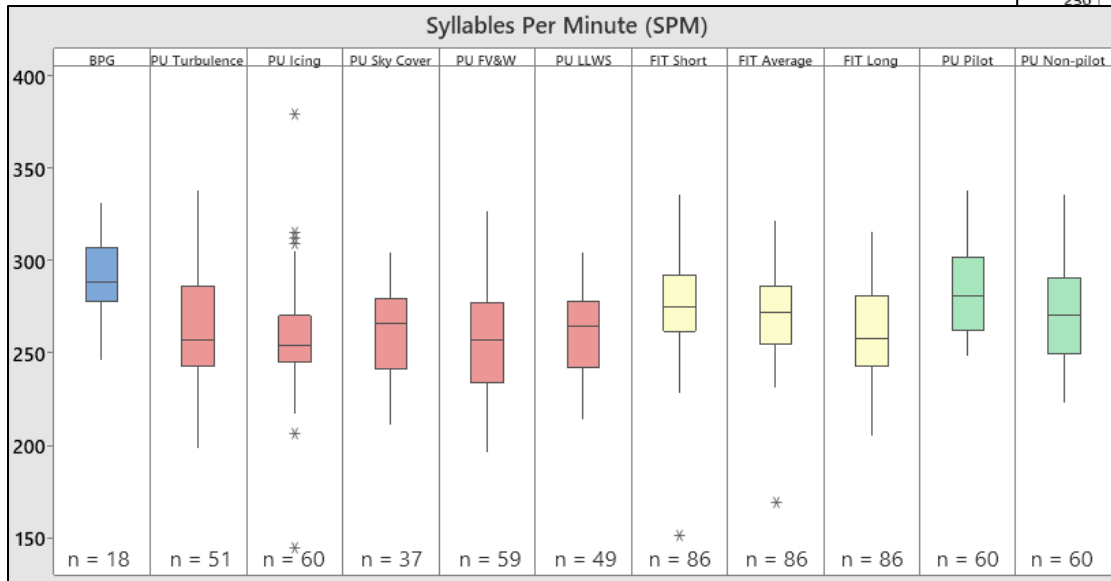
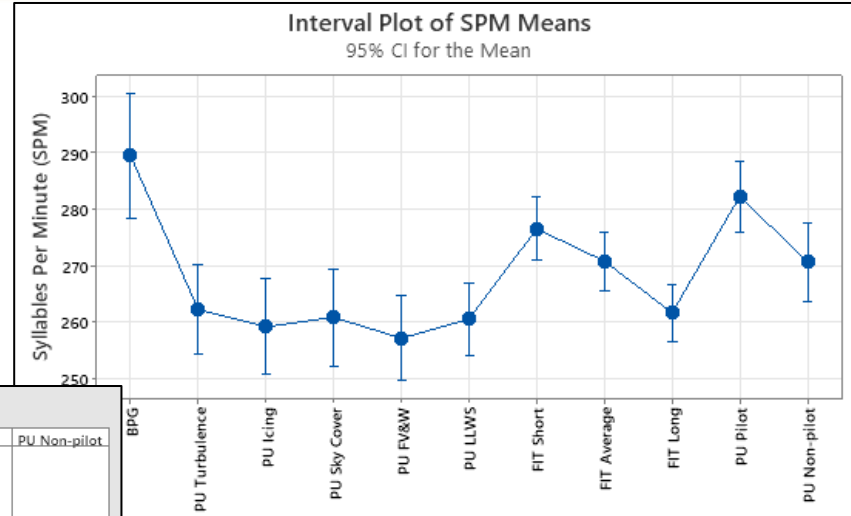
Words Per Minute (WPM)

- $$\frac{\text{Number of Words}}{\text{Phonation Time (Speaking Time) in minutes}}$$
- Indicates information production rate



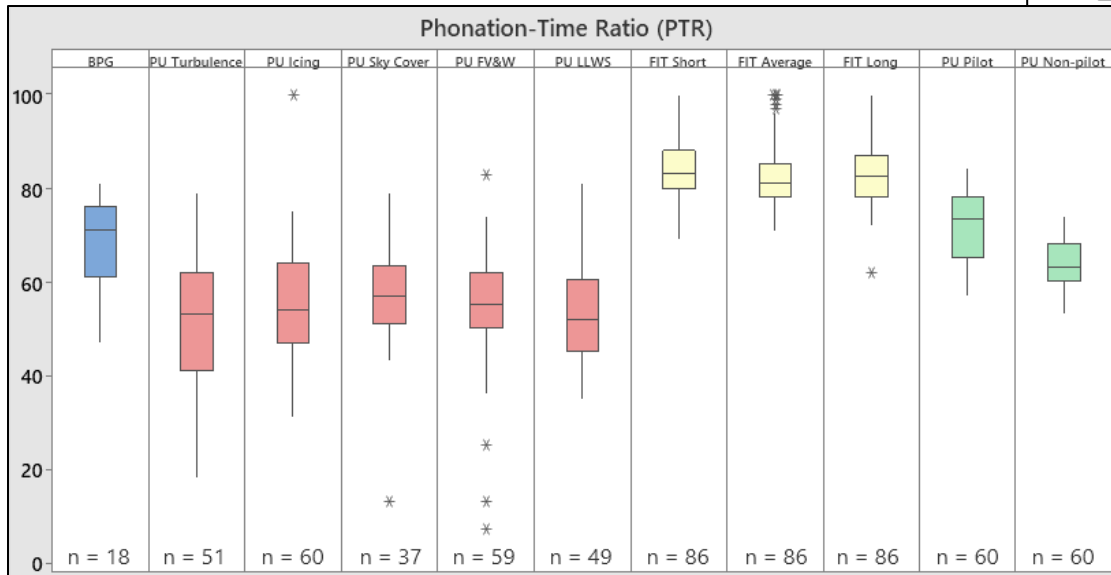
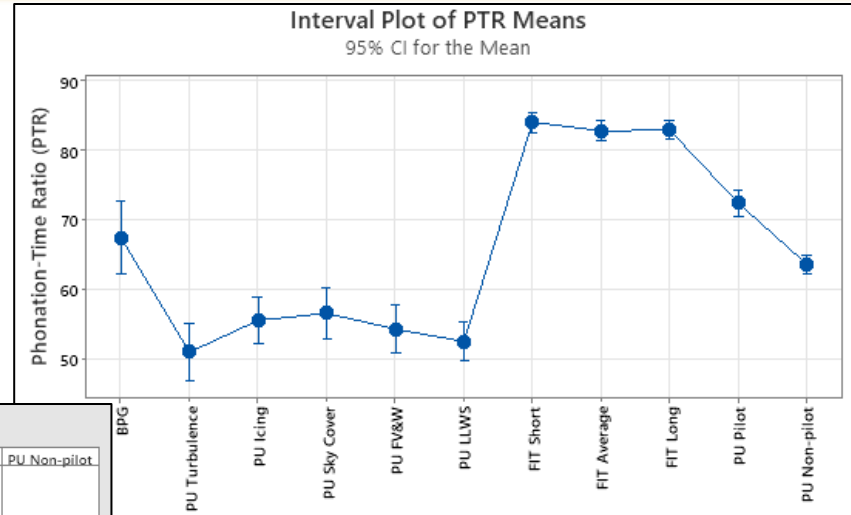
Syllables Per Minute (SPM)

- $$\frac{\text{Number of Syllables}}{\text{Phonation Time (Speaking Time) in minutes}}$$
- Indicates articulation rate



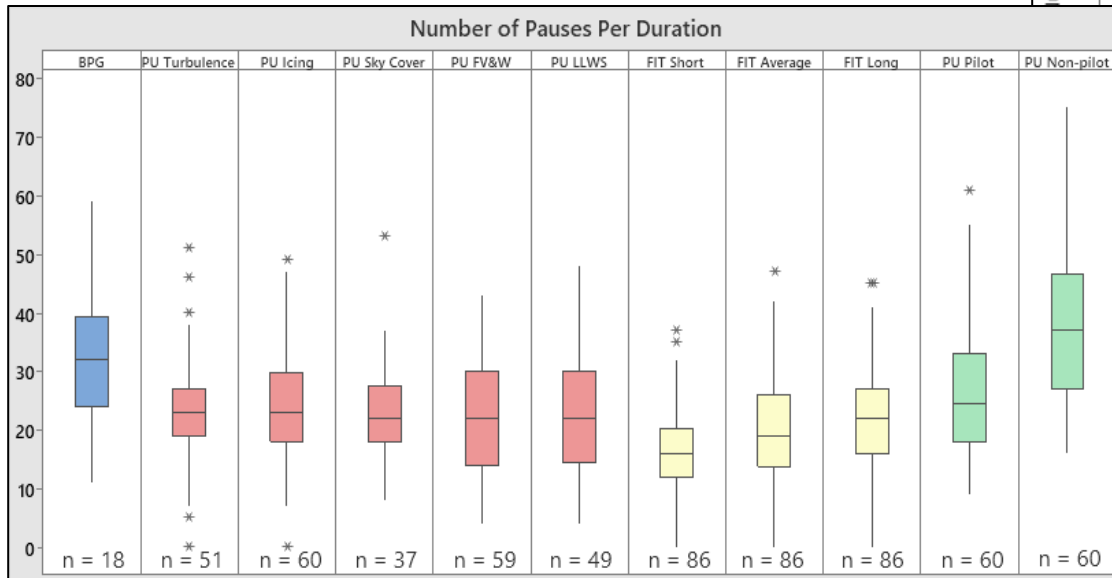
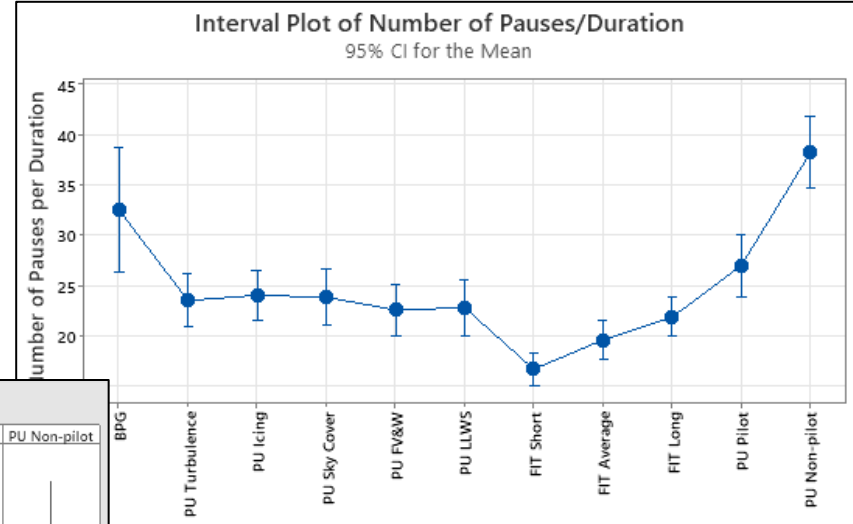
Phonation-Time Ratio

- $\frac{\text{Phonation Time (Speaking Time)}}{\text{Total Duration of audio}}$
- Indicates duty cycle



Number of pauses per duration

- $$\frac{\text{Number of Pauses}}{\text{Total Duration of audio}}$$
- Indicates speech density





Business Plan Goal Deliverable

WER vs Cadence Metrics: **BPG Data** (18 PIREPs)

Business Plan Goal (BPG)
Deliverable in March 2023

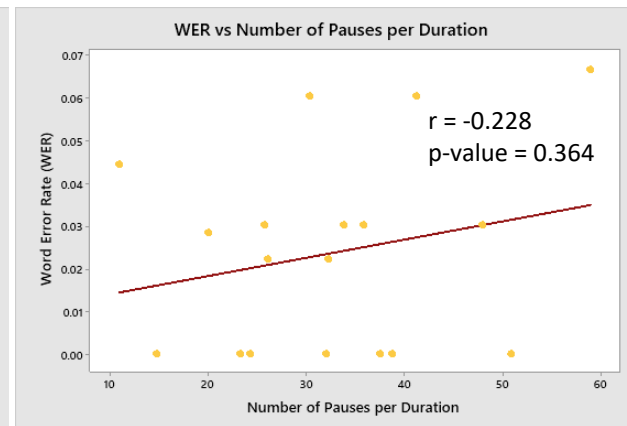
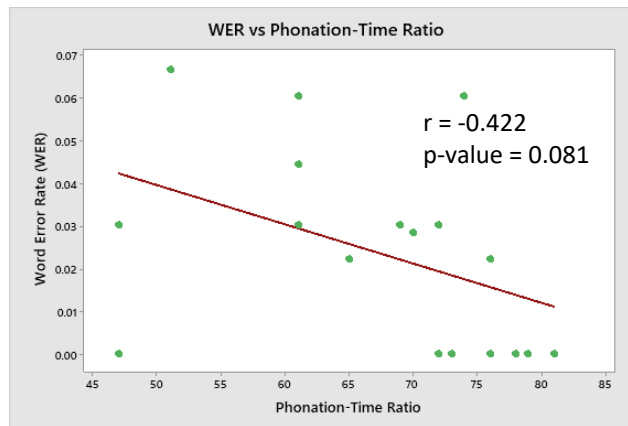
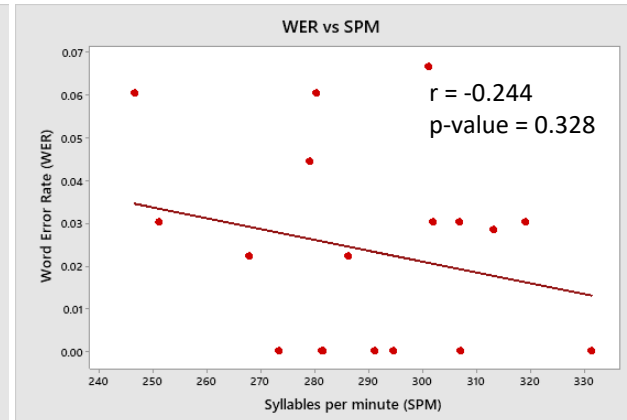
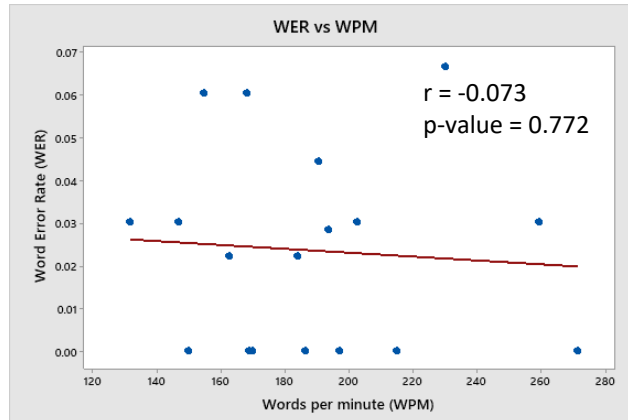
Initial study to explore Word Error
Rate (WER) vs Cadence metrics

18 audio PIREPs recorded by 6 team
members (read from scripts)

Summary of results:

Statistically non-significant
correlation

- WER decreases as speech rate increases (i.e., participants speak faster)
- WER decreases as speaking time per total duration increases (i.e., fewer pauses)





Business Plan Goal Deliverable

WER vs Cadence Metrics: **Purdue 282 PIREPs**

The WER studied against the four selected cadence metrics

The 282 PIREPs studied as a single sample

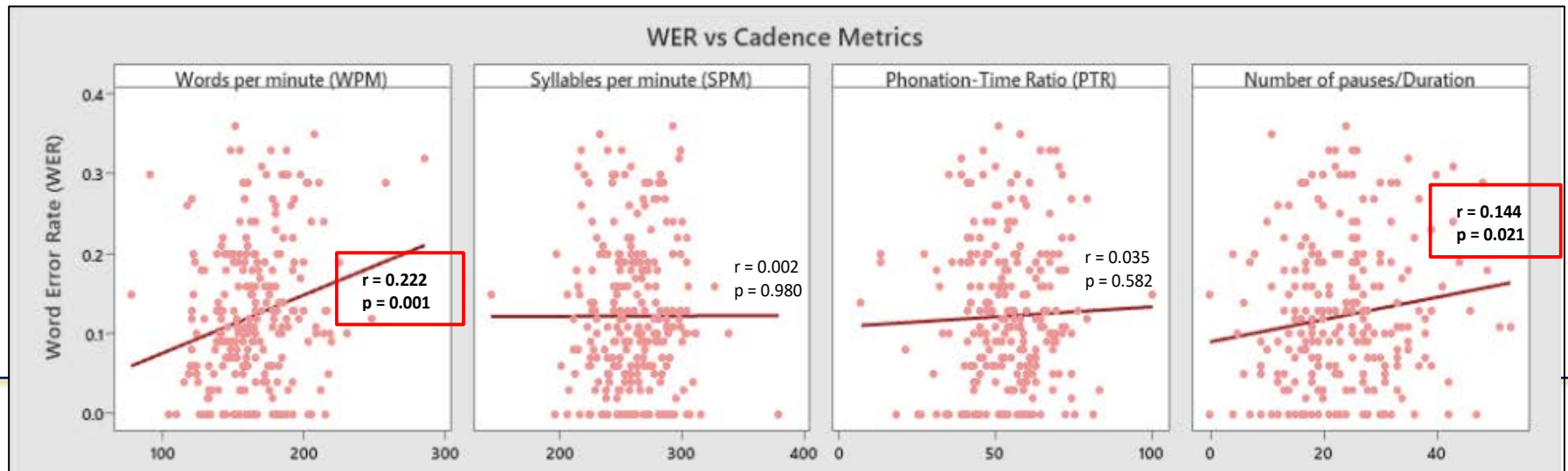
Summary of results:

Statistically significant **POSITIVE** correlation

- WER vs WPM – error increases as speech rate increases
- WER vs Number of Pauses/Duration – error increases as number of pauses increase

Statistically non-significant correlation

Pilot Demographics	Highlights from the survey responses
Age Category	51% respondents above the age of 51 22% respondents between 31 – 50
Base of operations	47% respondents from Florida, California, Texas, and Illinois Responses from 45 states plus D.C. and Puerto Rico
Flight Hours	48.2% respondents had less than 1000 flight hours as PIC 17% respondents had more than 5000 flight hours as PIC
Certifications	80% respondents held Private, Commercial or Instructor CFI Certification 5.1% respondents held Student, Sport or Recreational Certification.





Business Plan Goal Deliverable

WER vs Cadence Metrics: **Purdue 282 PIREPs**

The WER studied against the four selected cadence metrics

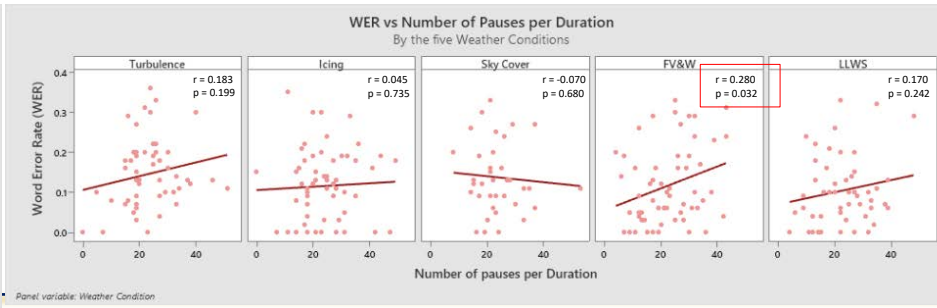
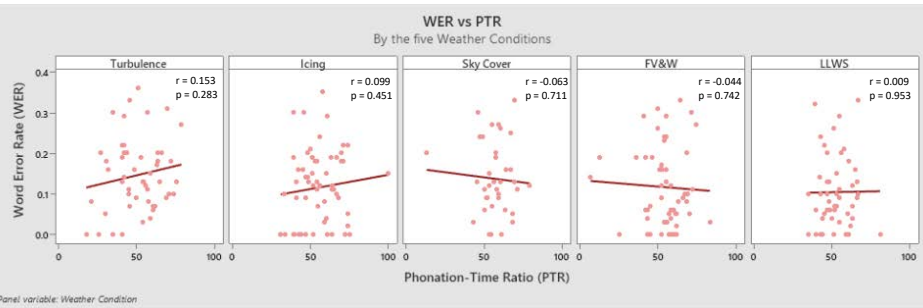
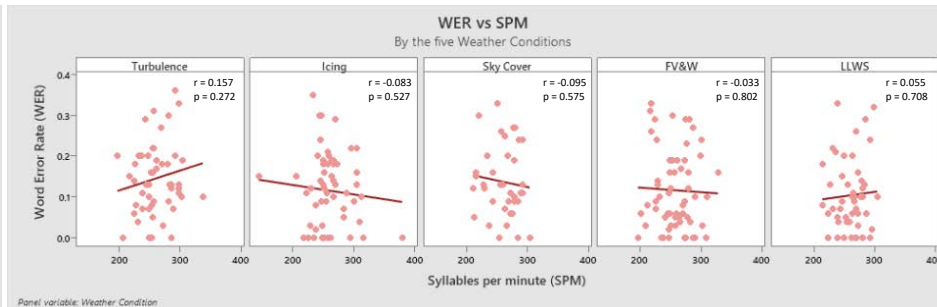
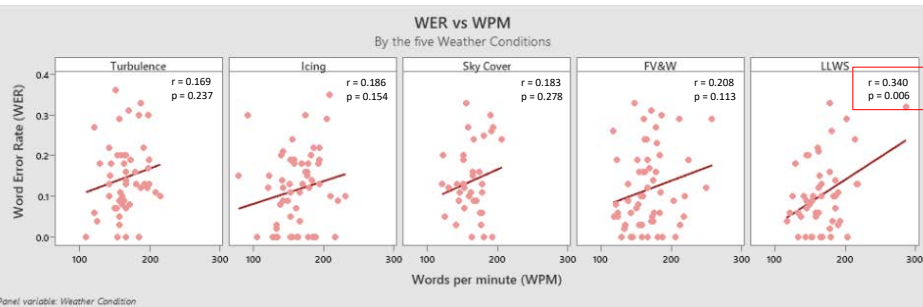
The 282 PIREPs studied across five weather conditions – Turbulence, Icing, Sky Cover, Flight Visibility, LLWS

Summary of results:

Statistically significant POSITIVE correlation

- WER vs WPM – LLWS-related PIREPs
- WER vs Number of Pauses/Duration – Flight Visibility and Weather (FV&W) PIREPs

Statistically non-significant correlations





Business Plan Goal Deliverable

WER vs Cadence Metrics: Purdue 282 PIREPs

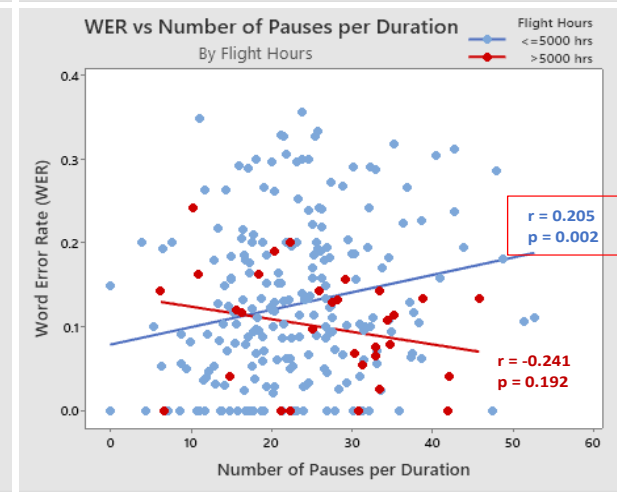
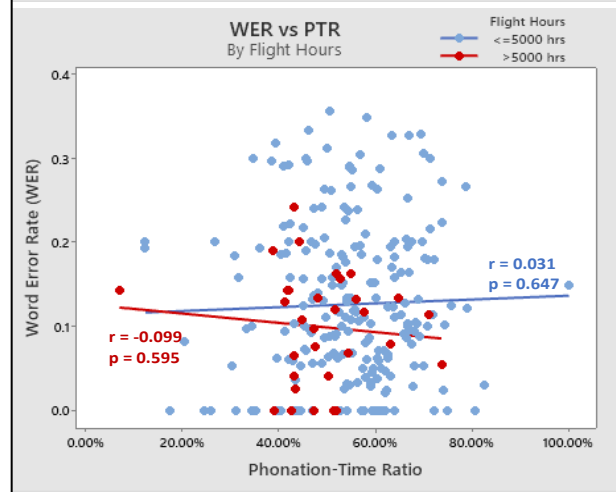
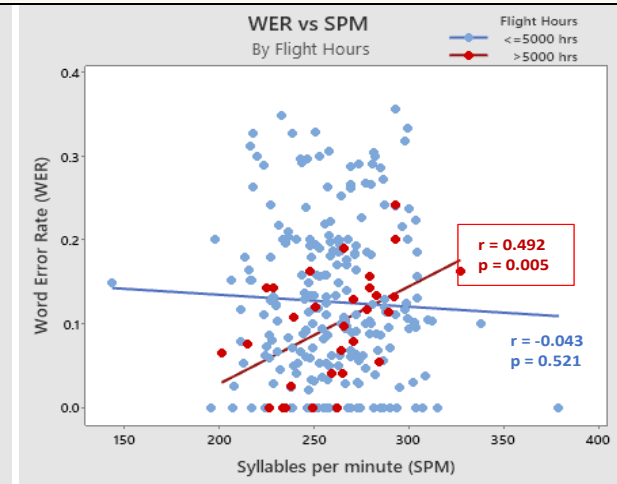
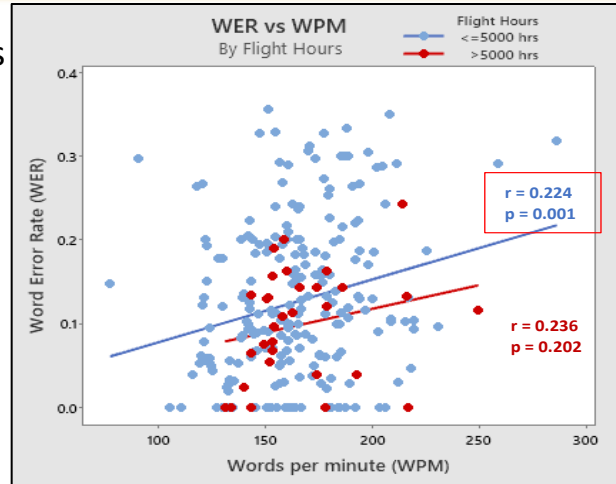
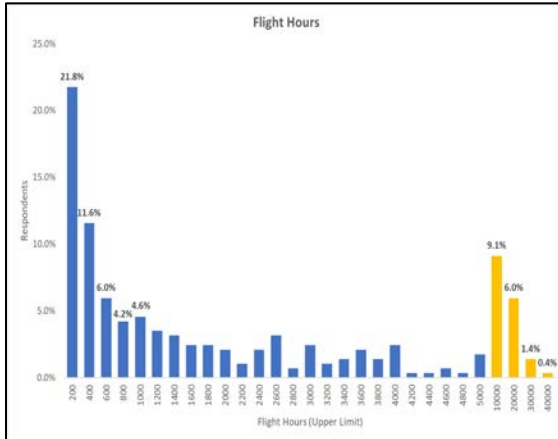
WER vs cadence metrics for two groups

- FH ≤ 5000
- FH > 5000

WER (FH ≤ 5000) > WER (FH > 5000)

Statistically significant correlations

- WER vs WPM: FH ≤ 5000
- WER vs SPM: FH > 5000
- WER vs Number of Pauses/Duration: FH ≤ 5000



WER vs Cadence Metrics: FIT 258 PIREPs

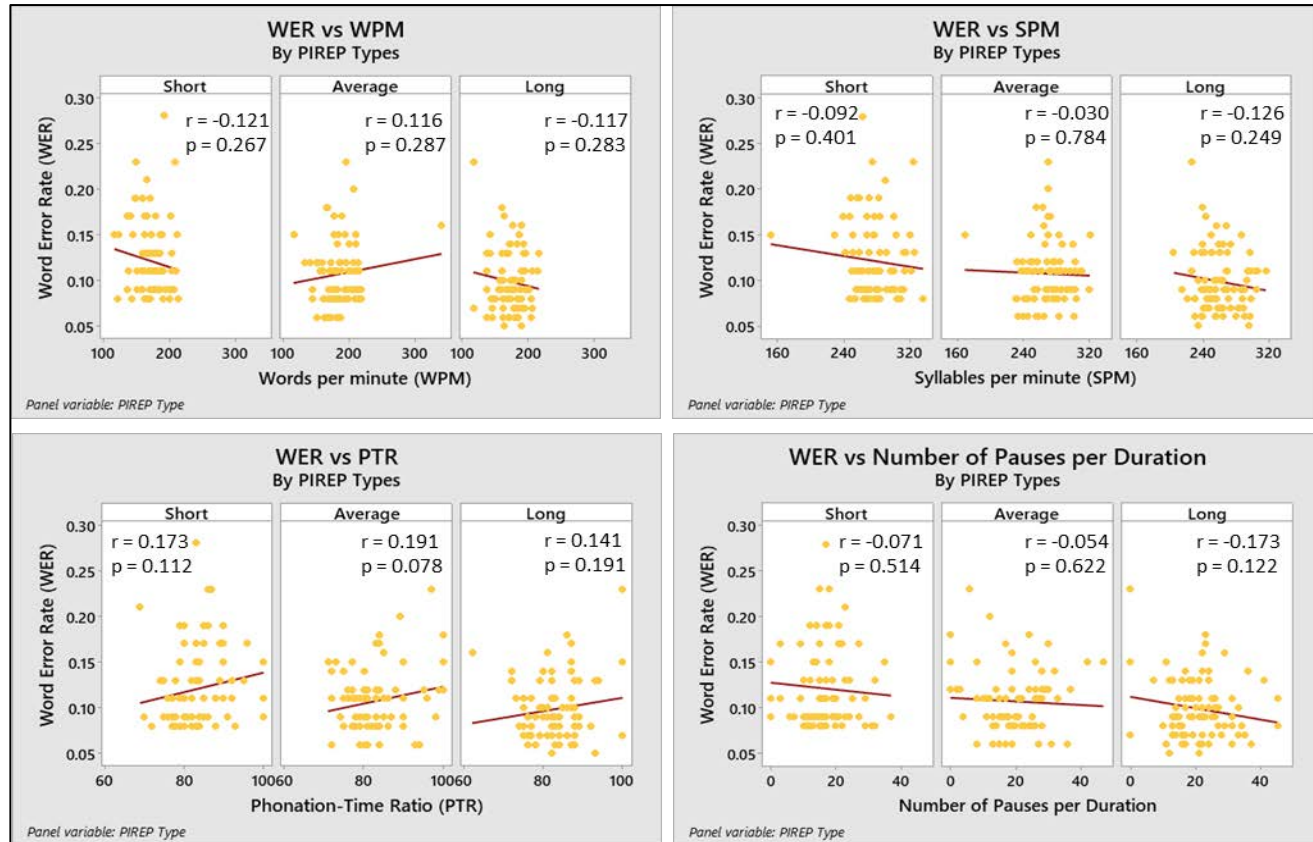
86 audio PIREPs each in Short, Average, and Long (read from scripts)

	Number of Words		
	Original Script	Transcription of audio PIREP	
		Average	Range
Short	46	52	46 – 56
Average	58	64	60 – 67
Long	81	90	86 – 94

Summary of results:

Statistically non-significant and weak correlation ($|r| < 0.25$)

- WER Short > Average > Long
- WER decreases as speech rate increases (i.e., participants speak faster)
- WER increases as speaking time per total duration increases (i.e., fewer pauses)





AWIP Pilot and Non-Pilot Data (120 PIREPs)

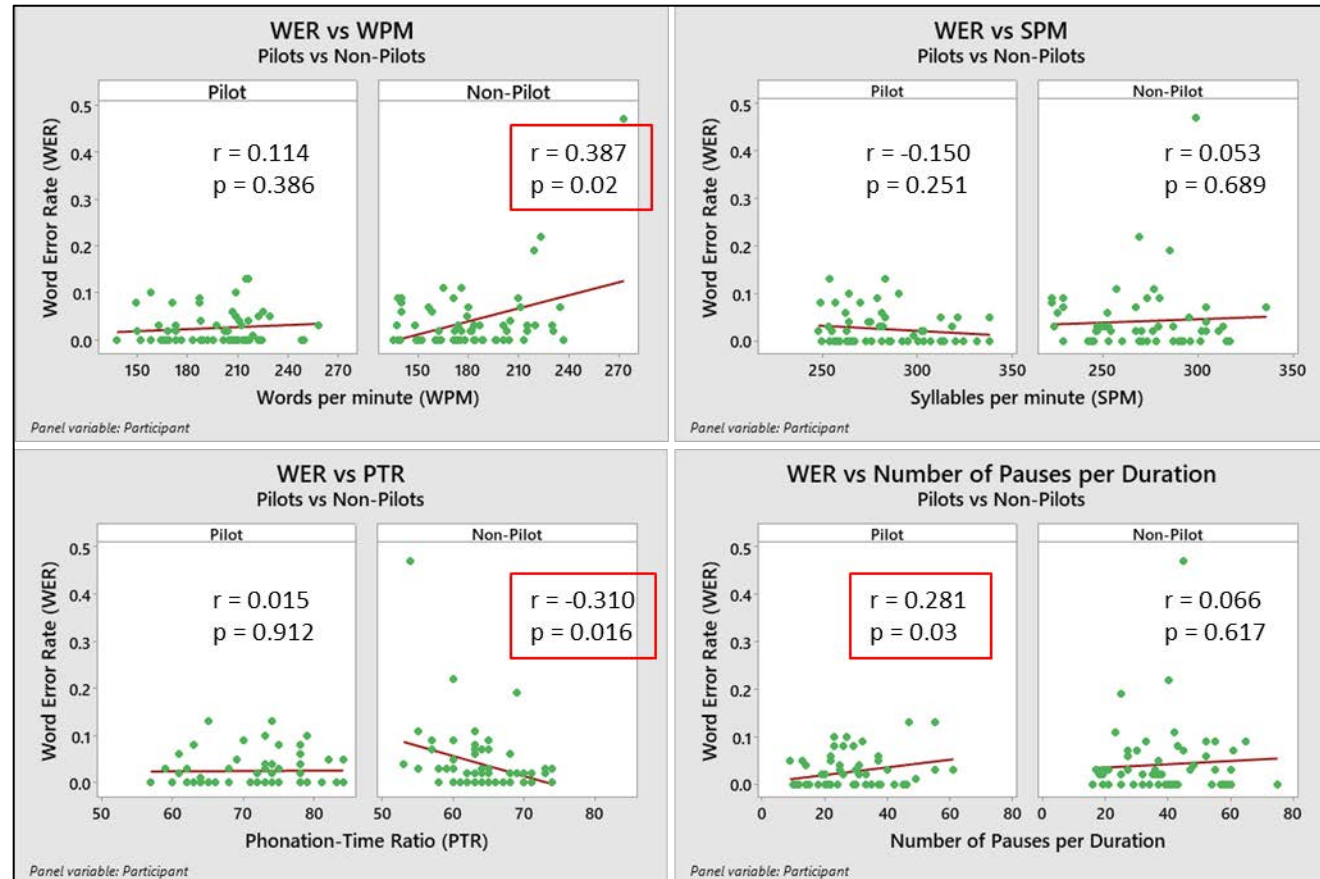
60 audio PIREPs each from Pilots and Non-pilots (read from scripts)

Pilots and Non-pilots read the same set of PIREP scripts

Summary of results:

Statistically non-significant weak correlation except in 3 cases

- WER Non-Pilot > Pilot
- WER increases as speech rate increases (i.e., participants speak faster) for non-pilots
- WER decreases as speaking time per total duration increases (i.e., fewer pauses)





Comparison of PIREP Submittal Tool Installed at both Purdue and FIT

Independent PIREP Submittal Tool running at FIT

- Transferred files and instructions from Purdue to FIT
- Installed required software and the tool
- Troubleshoot errors and version control
- Independent copy of the tool running at FIT since 04/05

Experiment to compare the performance of the PIREP submittal tool

- Run selected files at Purdue and FIT
- String comparison of PIREP transcription
- String comparison of PIREP codes

	PIREP Transcription Output			PIREP Codes Output		
	Tool at Purdue	Tool at FIT	String comparison score	Tool at Purdue	Tool at FIT	String comparison score
FIT Dataset						
Purdue Dataset						



Comparison of PIREP Submittal Tool Installed at both Purdue and FIT

Independent PIREP Submittal Tool running at FIT

No difference in PIREP transcriptions

No difference in PIREP codes

The tools are performing exactly the same

File (.wav)		PIREP Transcriptions			PIREP Codes		
		FIT Transcription	Purdue Transcription	Transcription String Comparison	FIT PIREP Code	Purdue PIREP Code	PIREP Code String Comparison
FIT 258 PIREPs Dataset	1S.wav	a routine pirep over delevingne five one two zulu at one thousand five hundred agl type of aircraft boeing seven thirty seven reporting broken ceiling at one thousand one hundred agl flight visibility is to statute miles with haze and smoke temperature is thirty degrees celsius	a routine pirep over delevingne five one two zulu at one thousand five hundred agl type of aircraft boeing seven thirty seven reporting broken ceiling at one thousand one hundred agl flight visibility is to statute miles with haze and smoke temperature is thirty degrees celsius	TRUE	UA/FL015/WX	UA/FL015/WX	TRUE
	1A.wav	a routine pirep ten miles southwest of appleton vor at one five one six sulu at eight thousand five hundred type of aircraft beach two hundred super king air reporting the base of a broken cloud layer is at six thousand five hundred flight visibility is three statute miles with haze and smoke temperature is twenty degrees celsius and light turbulence	a routine pirep ten miles southwest of appleton vor at one five one six sulu at eight thousand five hundred type of aircraft beach two hundred super king air reporting the base of a broken cloud layer is at six thousand five hundred flight visibility is three statute miles with haze and smoke temperature is twenty degrees celsius and light turbulence	TRUE	UA/OV ATW 225010/FL085/S K BKN/TB LGT CHOP	UA/OV ATW 225010/FL085/S K BKN/TB LGT CHOP	TRUE
	1L.wav	routine pirep sixty four nautical miles on the zero nine zero radial of oklahoma city v. o. r. at one five two two zulu at eight thousand msl type of aircraft cessna one seventy two reporting a scattered cloud layer with bases at nine thousand and tops unknown flight visibility is five statute miles with haze outside air temperature is minus four degrees celsius wind is two four five at four zero knots light turbulence sky is clear thunderstorms south moving east with occasional cloud to cloud light	routine pirep sixty four nautical miles on the zero nine zero radial of oklahoma city v. o. r. at one five two two zulu at eight thousand msl type of aircraft cessna one seventy two reporting a scattered cloud layer with bases at nine thousand and tops unknown flight visibility is five statute miles with haze outside air temperature is minus four degrees celsius wind is two four five at four zero knots light turbulence sky is clear thunderstorms south moving east with occasional cloud to cloud light	TRUE	UA/OV OKC 090004/FL080/S K SKC/TA M04/TB LGT CHOP	UA/OV OKC 090004/FL080/S K SKC/TA M04/TB LGT CHOP	TRUE
Purdue 120 PIREPs Dataset	NP1A1.wav	atc cessna one seventy two november one two three four charlie at hangtown hotel november whiskey three o. four radial eleven miles out of auburn municipal air is clear and stable no turbulence found where it was forecast earlier thank you	atc cessna one seventy two november one two three four charlie at hangtown hotel november whiskey three o. four radial eleven miles out of auburn municipal air is clear and stable no turbulence found where it was forecast earlier thank you	TRUE	UA/OV HNW 034011/FLUNK N/TP C172	UA/OV HNW 034011/FLUNK N/TP C172	TRUE
	P1B1.wav	approach piper cub one two three papa with pilot report piper cub one two three papa experience extreme turbulence fifteen miles southwest southwest of five bill at two thousand five hundred feet over	approach piper cub one two three papa with pilot report piper cub one two three papa experience extreme turbulence fifteen miles southwest southwest of five bill at two thousand five hundred feet over	TRUE	UUA/FL025/TP J3/TB EXT	UUA/FL025/TP J3/TB EXT	TRUE

Voice Analysis Software: Praat

- Speech analysis in phonetics and linguistics created by Boersma and Weenink (2022) of the University of Amsterdam.
- Provides a detailed acoustic analysis of speech recordings, including gender and mood.

Gender	Mood of Speech	Pitch Range (Hz)
Male	No emotion, normal	97-114
Male	Reading	115-135
Male	Speaking passionately	136-163
Female	No emotion, normal	164-197
Female	Reading	198-226
Female	Speaking passionately	227-245

- Dataset: 60 pilots and 60 non-pilots
- Categorized the results of the speech analysis into distinct mood categories for both pilots and non-pilots (i.e., 'no-emotion,' 'reading,' 'speaking passionately,' and 'voice not recognized') for each of the 120 audio files.
- Used pre-existing Word Error Rate (WER) data for each audio file, providing an additional layer of analysis to correlate moods with speech recognition accuracy.



Metrics related to Emotion Analysis

One-Tailed Distribution	P - NE	P - R	P - SP	NP - NE	NP - R	NP - SP	P - VNR	NP - VNR
P - NE	X							
P - R	0.2316	X						
P - SP	0.1391	0.2913	X					
NP - NE	0.3231	0.2692	0.3950	X				
NP - R	0.0664	0.3631	0.2360	0.2109	X			
NP - SP	0.0364	0.3288	0.0654	0.0593	0.1163	X		
P - VNR	0.3555	0.2654	0.3860	0.4847	0.2141	0.0589	X	
NP - VNR	0.4168	0.2216	0.1228	0.2828	0.0590	0.0333	0.3149	X



Metrics related to Emotion Analysis

Word Error Rate (WER) vs. Mood

Pilots	No-Emotion	Reading	Speak Passionately	Voice not recognized
Average	0.0179	0.0533	0.0277	0.0233
Max	0.0900	0.1300	0.1300	0.0400
Min	0	0	0	0
Standard Deviation	0.0271	0.0556	0.0370	0.0170
Variance	0.0007	0.0031	0.0014	0.0003
Number of Pilots /Samples	24	3	30	3

Non-Pilots	No-emotion	Reading	Speak Passionately	Voice not recognized
Average	0.0240	0.0372	0.0767	0.0160
Max	0.0900	0.2200	0.4700	0.0700
Min	0	0	0	0
Standard Deviation	0.0350	0.0548	0.1120	0.0211
Variance	0.0012	0.0030	0.0125	0.0004
Number of Pilots/Samples	10	25	15	10



Metrics related to Emotion Analysis

Word Error Rate (WER) vs. Mood

Non-Pilots	No-emotion	Reading	Speak Passionately	Voice not recognized
Average	0.0240	0.0372	0.0767	0.0160
Max	0.0900	0.2200	0.4700	0.0700
Min	0	0	0	0
Standard Deviation	0.0350	0.0548	0.1120	0.0211
Variance	0.0012	0.0030	0.0125	0.0004
Number of Pilots/Samples	10	25	15	10

- **Non-Pilots Speaking Passionately vs. Pilots No Emotion:**
 - With a p-value of 0.0364, there is an indication that non-pilots may be more expressive in their passionate speech compared to the emotionally neutral speech of pilots.
 - This finding may be useful for further analysis because “no emotion” classified speech was also associated with lower WER.
- **Non-Pilots - Voice Not Recognized (NP – VNR) vs. Non-Pilots - Speaking Passionately (NP - SP):**
 - With a p-value of 0.0333, this comparison shows with statistical significance that when non-pilots' voices are not recognized, their expressiveness while speaking passionately is distinctly different from those times when it is not recognized.

- Prototype evaluated on how well it could receive transcriptions and convert them into coded PIREPs.
- This research focused on the 60 pilot and non-pilot Purdue audio recordings.
 - Training of the prototype did not include the FIT audio recordings.
- The objective was to discern the essential components of PIREPS and articulate clear expectations for the prototype performance.

- 1. Addition:** This occurs when an extraneous entry was added.
Example: No report of turbulence in the scenario (/TB Reported a MOD CHOP).
- 2. Substitution:** This occurs when a segment of an entry was replaced.
Example: We are at 8500 feet (/FL Reported as 005 or UNKN).
- 3. Partial Deletion:** This occurs when a segment of a correctly recognized PIREP element entry is deleted.
Example: Scattered clouds cover with bases at 7000 feet and tops at approximately 8500 feet. (/SK Reported as SCT).
- 4. Complete Deletion:** This occurs when the prototype expunges the entire entry, resulting in non-recognition.
Example: This is Beech Baron November 123 Tango (/TP Reported as).

Accurate recognition vs. total expected
(counts and total percentages)

Mandatory PIREP Elements				
PIREP Element	Accurate Recognition (Counts)	Out of	Total Expected (Counts)	Total Percentage
UA / UUA	55		60	91.7 %
OV	24		48	50 %
TM	13		27	48.1 %
TP	27		60	45 %
FL	17		55	30.9 %

The highest level of “**Accurate Recognition**” at **91.7%** for the **UA / UUA**, and the lowest level of accurate recognition at **30.9%** for **FL** category.

Accurate recognition vs. total expected (counts and total percentages)

Additional PIREP Elements				
PIREP Element	Accurate Recognition (Counts)	Out of	Total Expected (Counts)	Total Percentage
TB	21		34	61.8 %
WX	5		10	50 %
TA	10		26	38.5 %
IC	8		27	29.6 %
SK	3		23	13 %
WV	0		5	0 %
RM	0		4	0 %

The highest level of “**Accurate Recognition**” at **61.8%** for the **TB** and the lowest level of accurate recognition at **0%** for **WV** and **RM** categories.

Technological improvements and training to reduce errors

- Speech-to-PIREP tool demonstrated robustness when processing spoken PIREPs with speech and cadence variations, of different lengths, and from pilots and non-pilot participants
- Technological improvements and further training is needed to reduce the inconsistencies and make the tool as close to error-free as possible

Suggestions related to speech cadence and how error rates may be reduced

- Experiments indicated that word error rate (WER) was higher for faster speakers, less phonation-dense audio, and shorter PIREPs
- Pilots may be advised to communicate PIREPs at a slower or moderate pace, avoid long pauses or higher number of pauses, and communicate concise but complete PIREP

Future work related to error propagation, accents, flight experience, stress, and background noise

- Experiments in AWIP.X.1 and AWIP.X.2 focused on the error rate of System 1 and how it was affected by the variations in speech cadence
- Further research is needed to understand the error propagation in Systems 2 and 3
- Impact of pilot accent, stress level, and background noise must be evaluated along with speech variations

Deliverable AWIP.X.1: Metric(s) to quantify cadence of PIREP submissions as affected by experience, stress, and environmental conditions

- ✓ Study cadence metrics against Word Error Rate (WER) of the PIREP tool
 - ✓ Exploratory analysis
 - ✓ Regression modelling
 - ✓ Other data science approaches
 - ✓ Additional experiments with age, flight hours
- ✓ Deliverable Reports

Deliverable AWIP.X.2: Comparison between remotely submitted (from AWIP.3.2), pre-existing (previously recorded from FAA or other sources), and in-person PIREP datasets to verify ML performance

- ✓ Design experiment to compare the performance of the PIREP submittal tool
 - ✓ Run selected files at Purdue and FIT
 - ✓ String comparison of PIREP transcription
 - ✓ String comparison of PIREP codes
- ✓ Deliverable Reports

Deliverable AWIP.X.3: Report of experimental results and completion of MinWxSvc recommendation



Thank you for your time! Questions?

Barrett Caldwell, PhD
Mary E. Johnson, PhD
Brandon J. Pitts, PhD
Shantanu Gupta

Purdue University

Deborah S. Carstens, PhD
Michael Splitt, MS
Faraz Amiraslani
Joshua Breininger
Rusheel Raj Panakadan

Florida Institute of Technology

Project Update, June 2024