Naval Surface Warfare Center,
Dahlgren Division

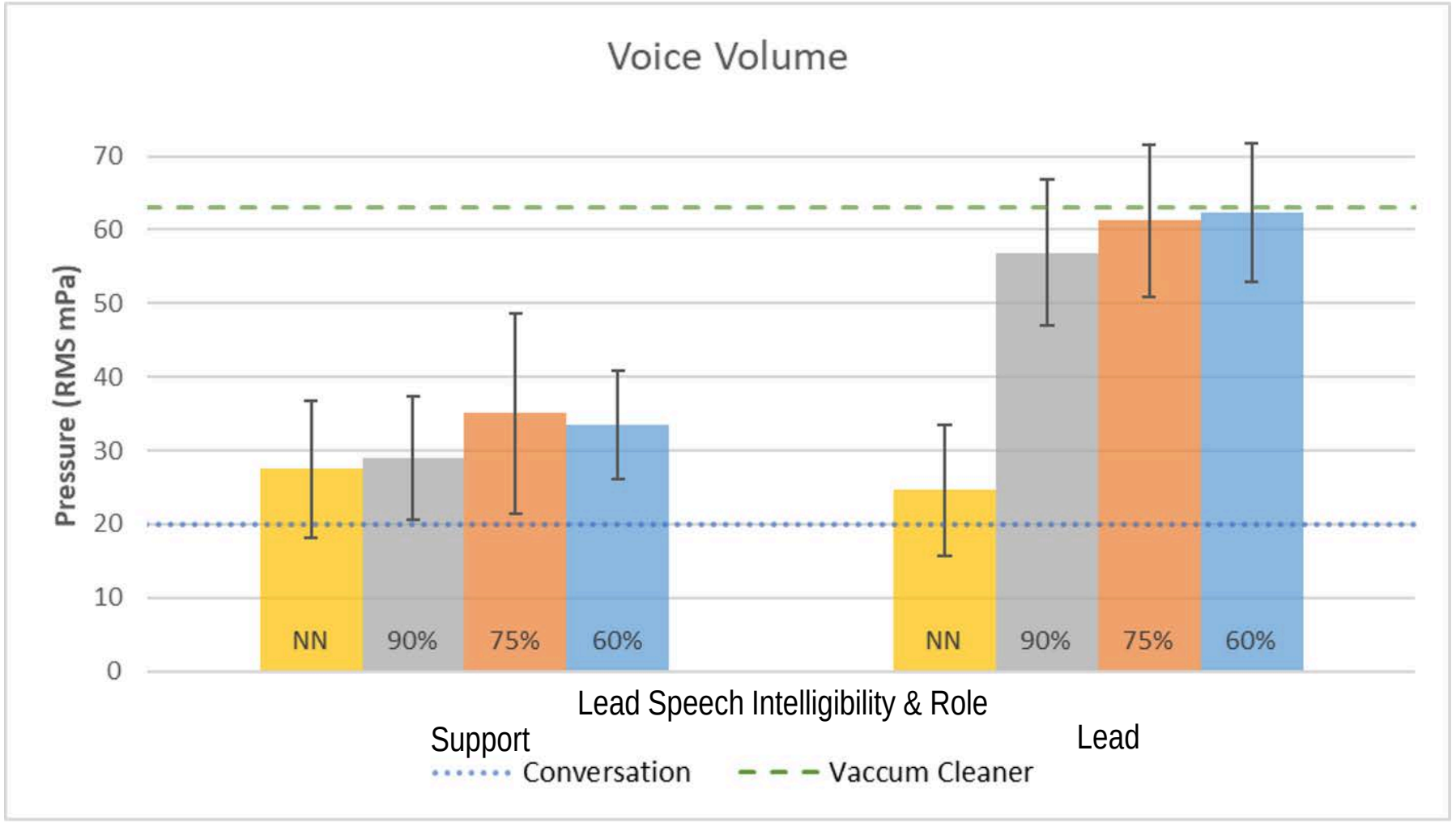
Performance in Noise (PiN)

2/9/2021

Background

- Background
  - Military tasks require rapid and reliable vocal communication.
  - Tasks often take place in environments with continuous and unexpected noise disrupting. (CIC, flight deck, engine room, etc.)
- History of PiN
  - Little data was available to quantify effects of noise in environment or Hearing Loss for tactical situations.
  - PiN sought to quantify communication noise impact on human performance in a tactical situation.
  - In noisy environments performance does suffer. Sailors speak louder, speak slower, communicate less accurately, share less information, and are less focused on visual information.

Example CIC



Problem

- Cause
  - Noisy environments
  - Loud environments
  - Poor communications equipment
- Effect Level 1
  - Reduced ability to vocally communicate
  - Reduced ability to perform tasks
    - Task time, task accuracy, attention, stress, frustration, yelling, mal-adaptive behaviors, idle time, situation awareness (SA), engagement level, engagement time, alternative communication time, teammate effects, etc.
  - Long term hearing damage
- Effect Level 2
  - Risks mission
  - Risks life

Current mitigations

- Wear hearing protection
  - Further reduced ability to vocally communicate and to perform tasks
- Neglect hearing protection
  - Further reduces ability to vocally communicate over time
- Active hearing protection
  - High cost, fear of damage, not used
- Increased yelling and asking for repeats
  - Further increases environment noise amplitude, causes delays
- Use text chat, switch station, or move closer to speaker
  - Reduces engagement with tactical system and does not gain full SA
  - Increases idle time waiting for response
  - Reduces teammate engagement with tactical system and reduces SA

Possible Solutions

- Promising
  - Provide augmented 2 way communication aid
- Not promising
  - Reduce environment noise
  - Reduce communications noise
  - Provide other modes of 2 way communication
  - Provide other modes of 1 way communication

Aims

- Develop an augmented 2 way communications aid
  - Deciphers vocal communications for noisy environments
  - Deciphers vocal communications for Navy jargon
    (abbreviations, phonetics, numbers, alternate meaning/structure/grammar)
  - Presents easily digestible communication

Goals

- Phase 1
  - Evaluate Speech to Text (STT) Commercial off the Shelf (COTS) models
- Phase 2
  - Train for specific real Navy jargon
  - Train for specific synthetic Navy noisy environments
- Phase 3
  - Human performance testing
- Phase 4
  - Generalize for real Navy jargon and real Navy noisy environments
- Phase 5
  - Transition

Figures of Merit

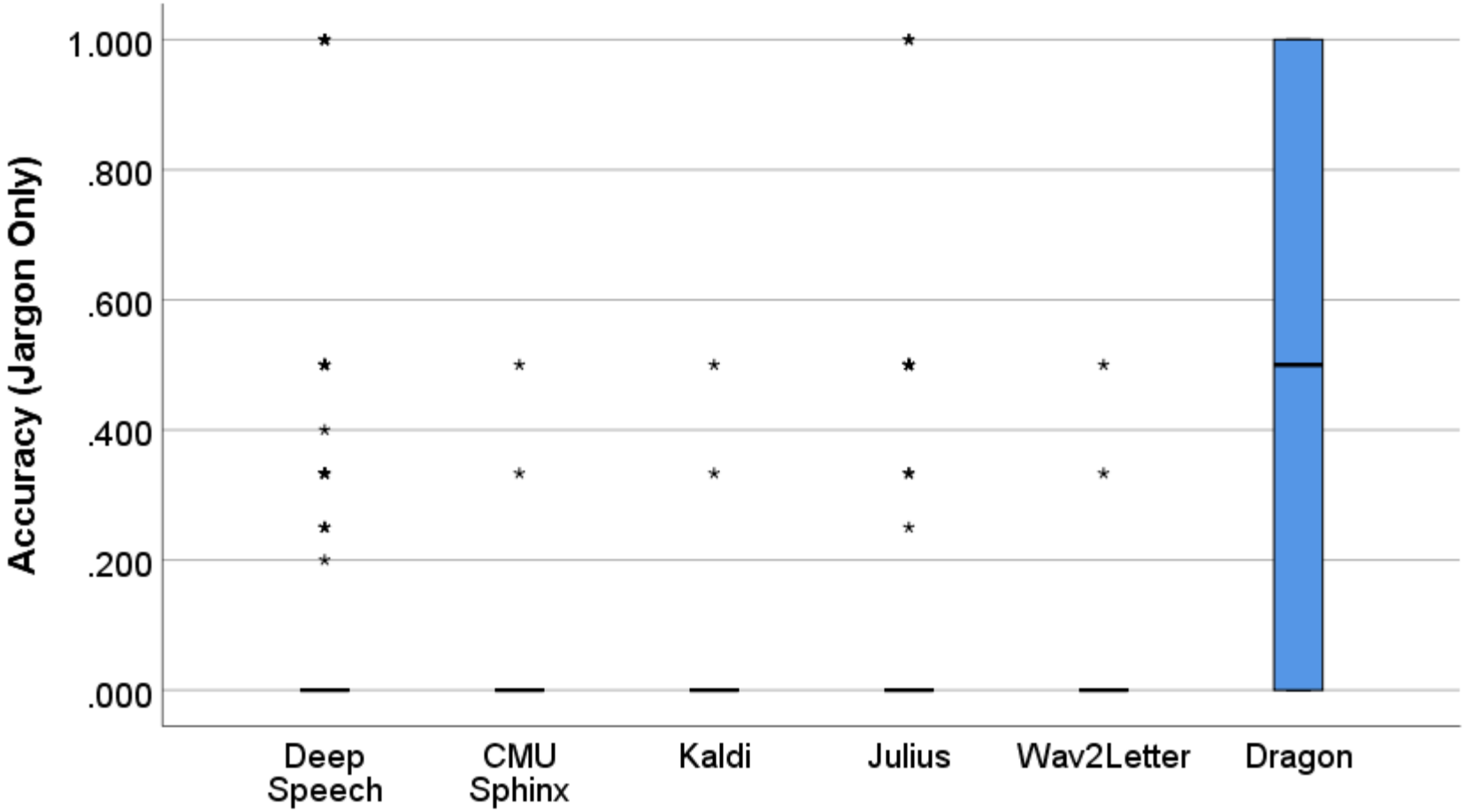
- Phase 1-2, 4: Select, train, evaluate
  - Word Error Rate (WER)
  - Jargon Accuracy Rate (JAR)
  - F1 Score
  - System resource usage
- Phase 3: Prototype aid usability & performance
  - Task Load Index (TLX)
  - Systems Usability Scale (SUS)
  - Engagement
  - Task time
  - SA focus time
  - Communications mode changes
  - Teammate forced backup
  - Eye movements
  - Task accuracy
  - Behavior changes

Methods

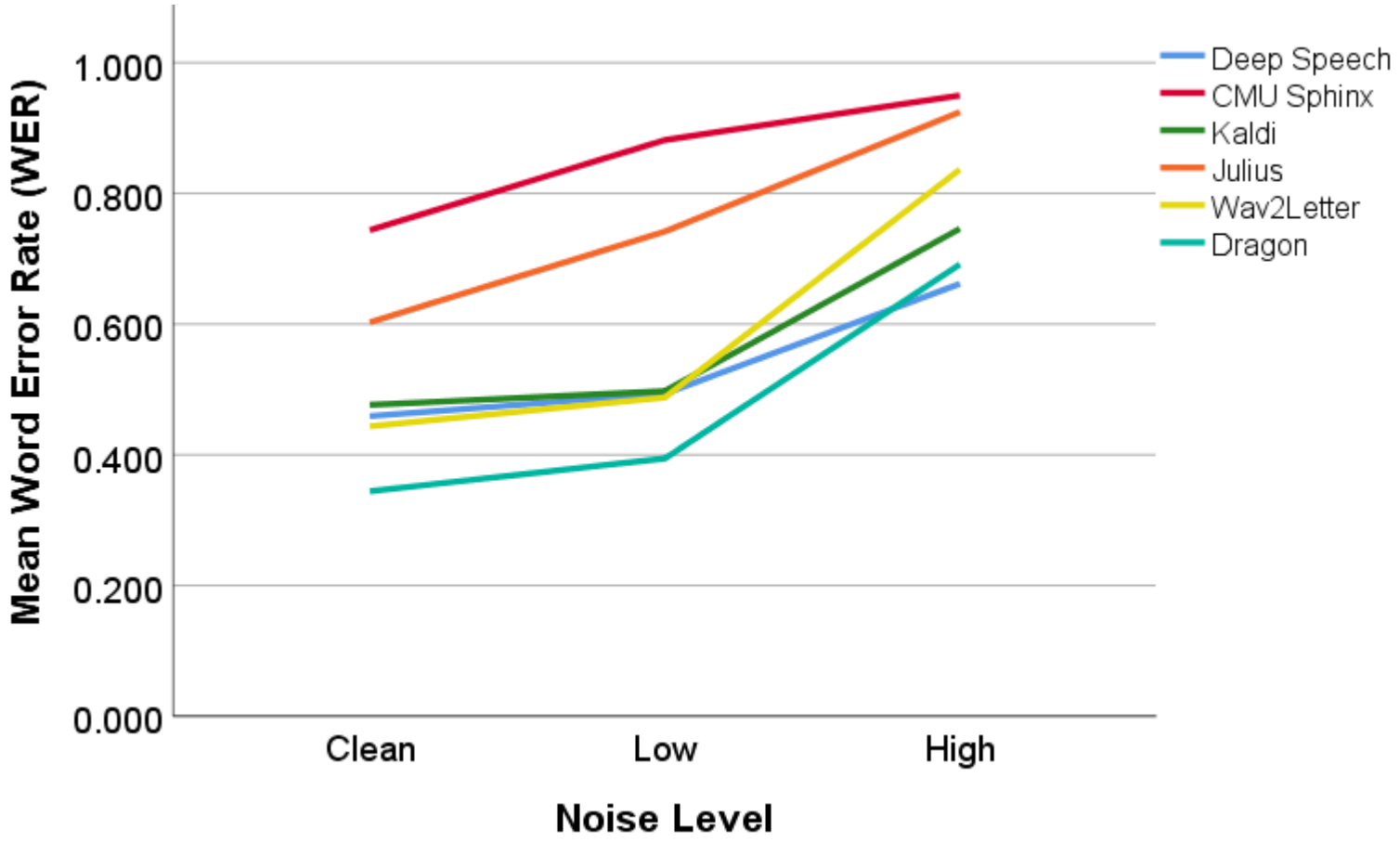
- **Gather Data** (record, transcribe, identify jargon)
- **Preprocess** (filter, sample rate, sample depth)
- **Prepare** (mix noise, slice utterances, summary stats)
- **Test** (identify metrics, batch process, analyze)

STT	X	Noise Amplitude	X	Noise Type
Deep-Speech (RNN)		None		White
CMU Sphinx (HMM)		Low		Gray
Kaldi (DNN-HMM & GMM-HMM)		High		Pink
Julius (HMM & DNN)				Engine room
Wav2letter (RNN & CNN)				
Dragon (DNN)				

Results: Jargon



Results: Noise



Overall Performance

STT Tool	Jargon Rank*	Performance Rank	Low Noise Rank	High Noise Rank	Cumulative Rank**
Dragon	1	1	1	2	5
Deep Speech	4	3	3	1	11
Wav2Letter	4	2	2	4	12
Kaldi	4	4	4	3	15
Julius	4	5	5	5	19
CMU Sphinx	4	6	6	6	22

Discussion

- Jargon (Initialisms – letter names)
  - Dragon out performed all others because it can interpret Initialisms based on letter names
- Jargon (Acronyms)
  - All tools struggled with inconsistent results due to unique Navy acronyms
- Noise
  - Deep Speech performed the best in the high noise conditions
  - Dragon suffered to cope with intelligibility issues driven by high noise
  - Deep Speech performed similarly to several others in the low noise conditions
  - Different noises had varying effects at low and high intensities
  - White & pink high noise had the greatest adverse effect on STT tool performance
  - Machine low noise had the greatest adverse effect on STT tool performance
- Selected for continued development
  - Dragon: Highest overall performance (all words & jargon) in clean and low noise conditions
  - Deep Speech: Most resilient to noise

Limitations

- Limited Jargon Set
  - Limited use scope of words
  - Initialisms with NATO phonetic alphabet were not fully examined
  - Initialisms with NATO phonetic numbers were not fully examined
  - Shortenings were not fully examined
  - Contractions were not fully examined
  - Alternate meaning/structure/grammar was not fully examined
- Floor Effects
- Ideal Microphone Performance
- Generic Machine Noise Profile

Future

- Development
  - Early results are promising and provide motivation to continue development
  - Training for specific real Navy jargon
  - Training for specific synthetic Navy noisy environments
  - Human performance testing
  - Incorporating fixes to existing limitations
- Impacts
  - Warfighter communications improvements
  - Warfighter task performance improvements
- Spin off
  - Warfighter use of voice controlled systems

Acknowledgements

- Office of Navel Research, Code 34
- Sonalysts