Read Back Error Detection using Automatic Speech Recognition

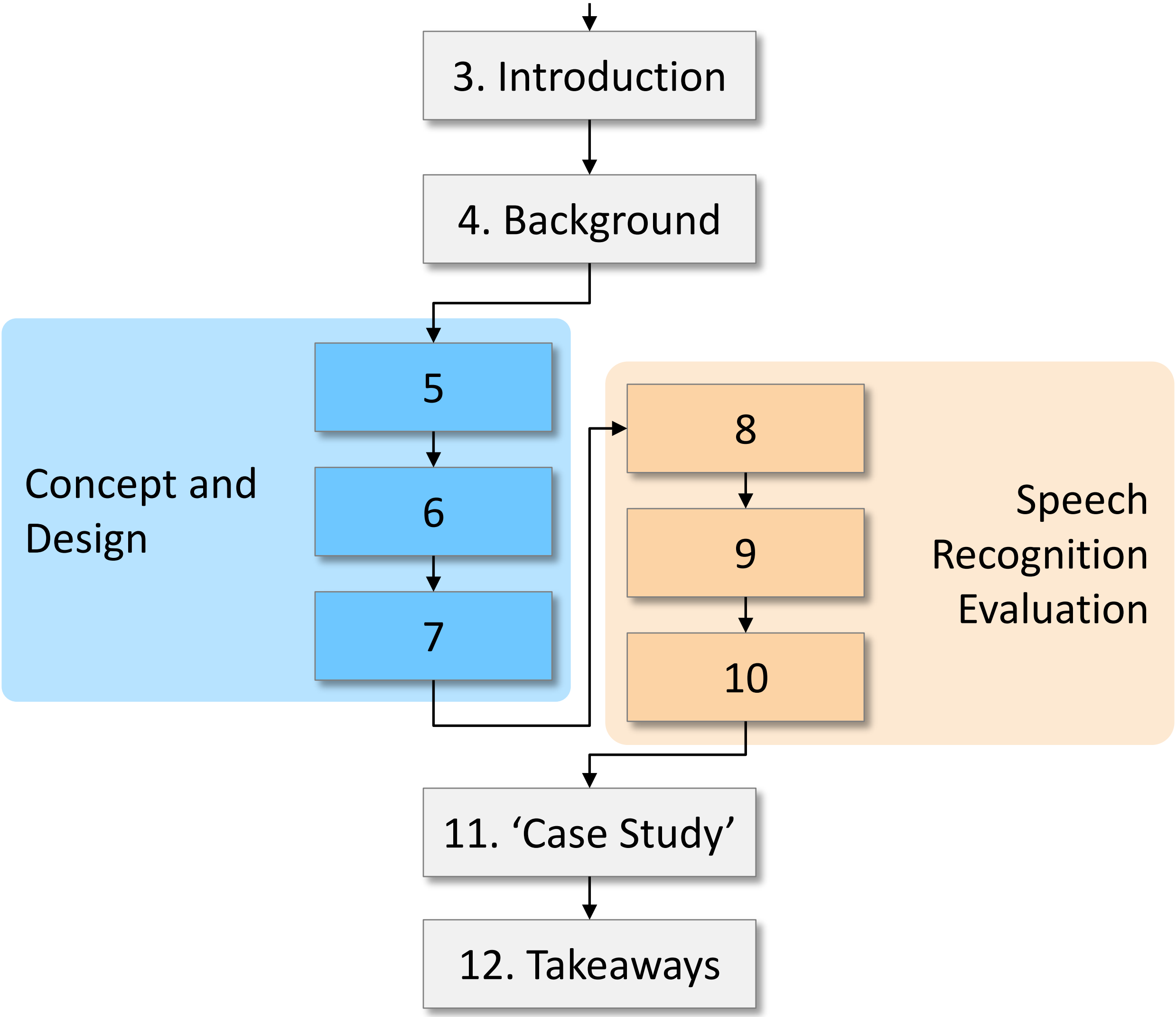
Shuo Chen
Hunter Kopald
Ronald S. Chong
Dr. Yuan-Jun Wei
Zachary Levonian

June 27, 2017

Approved for Public Release; Distribution Unlimited. Case Number 17-2100

MITRE

Outline

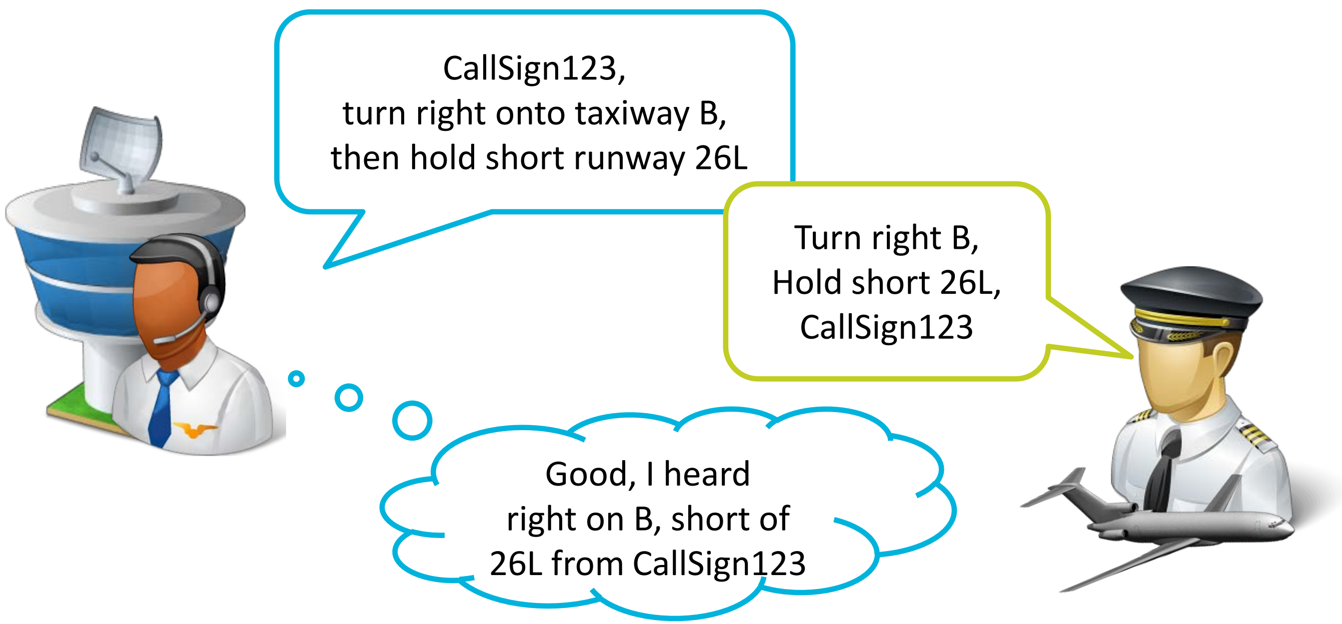


MITRE

Introduction

After a controller issues a clearance or instruction to a pilot, he or she must then "ensure that items read back are correct, ensure the read back of hold short instructions…, and ensure pilots use call signs and/or registration numbers in any read back acknowledging an air traffic clearance or ATC instruction"

– FAA Order JO 7110.65W



MITRE

Background

Automatic read back error detection is a popular idea with both current and former air traffic controllers and managers

The numbers on read back/hear back errors:

In the 1990's, Cardosi *et al.* reported
- ≤ 1% read back error rates in the tower domain
- 40% of read back errors were uncorrected
- **On average, one read back error per hour per frequency across all domains**

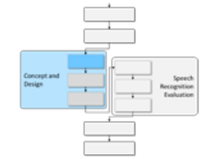
Eurocontrol found read back hear back errors were the most common controller-pilot communication problem

More recent study by Prinzo *et al.* reported
- 6% read back error rates in the TRACON domain
- 92% hear back error rate

Kopald and Goring found hear back errors accounted for 10.7% of runway incursions in a 6-year runway incursion dataset – **129 runway incursions in 6 years**

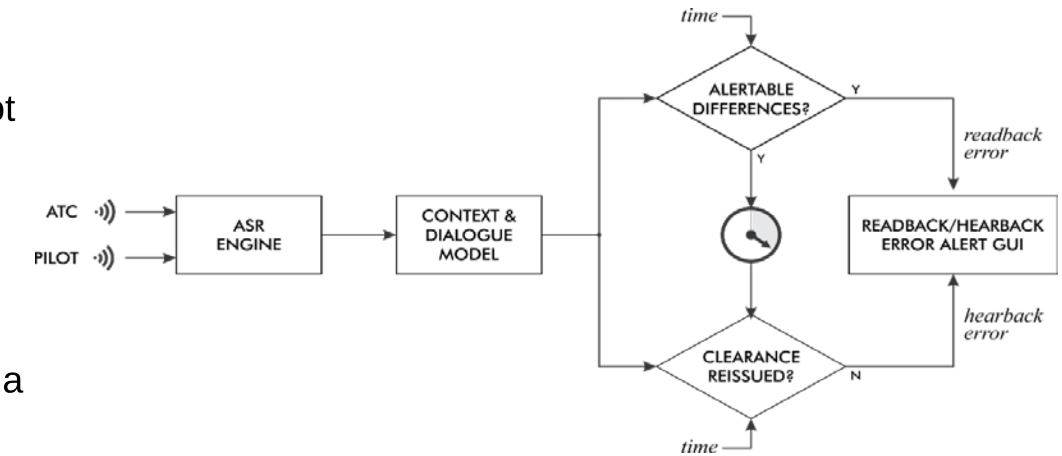
The risk of a runway incursion caused by a hear back error is very small, but the potential consequences of a single incursion may still warrant capabilities to mitigate this risk

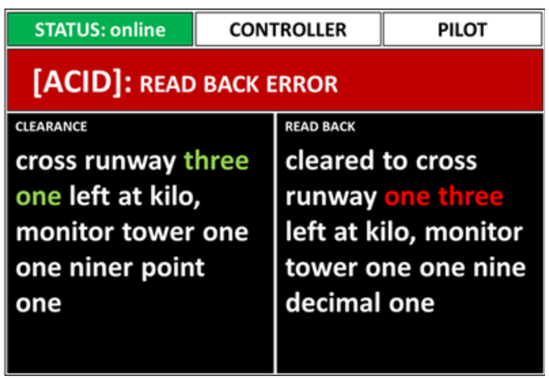
Concept and Design



A simple and intuitive design:

1. Passively monitor controller and pilot communications over the radio
2. Match transmission pairs, by ACID recognition and/or dialogue model
3. Extract commands for comparison
4. Indicate when a read back contains a mismatch or is absent altogether



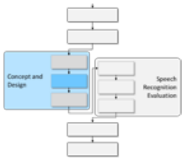


Designed a graphical user interface to satisfy the need for the following basic information, identified by subject matter experts:

- Aircraft identifier
- Indication that a read back error has occurred
- What the discrepancy was
- An audible alert if the read back error is not corrected

Key decision for the automation: is the read back error worth alerting?

MITRE



Alert Rules

Not all read back errors are created equal!

An analysis of 150 dialogues between controller and pilots at John F. Kennedy Airport (JFK) highlighted many interesting nuances…

[CALLSIGN], kennedy ground, 31L KE, taxi right A, hold short of J.

And A, short of J, for 31L KE, [CALLSIGN].

[CALLSIGN-0] at K, follow [CALLSIGN-1] off the right, cross runway 31L, and monitor tower 19.1.

Okay, behind the [CALLSIGN-1] from the right, follow her down K for 04L, and 19 1 to monitor, [CALLSIGN-0]

[CALLSIGN], cross runway 31L at K, and monitor tower 19.1.

Roger, cleared to cross runway 13L at K, and then monitor tower 119.1, [CALLSIGN].

[CALLSIGN], cross runway 31L at K, and monitor tower 19.1.

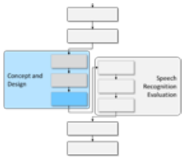
[Uh] Roger, cross runway 13L at K, and then monitor tower 1191, [CALLSIGN].

[CALLSIGN], it's runway 31L at K, cross runway 31L at K.

Sorry, runway 31L at K then, my apologies, [CALLSIGN], thank you

MITRE

Alert Rules (continued)



Summarizing observations and conclusions from the dialogue analysis:

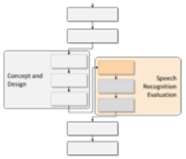
- **Order of elements** in read back does not need to match the original clearance ordering
- **Transfer of control/frequency**: controllers do not require facility name and frequency in the read back, and an acknowledgement of the transfer may suffice
- **Callsigns**: pilots almost always used their full callsign but controllers do not require a full or partial callsign when the identity of the speaker is unambiguous
- **Taxi commands**: pilots regularly dropped the left/right turn specification
- **Follow**: pilots almost always acknowledged the "follow" instruction with something semantically equivalent (e.g., "after", "behind").
- **Crossing instructions**: controllers almost always used the form "cross <runway> at <intersection>" and pilots always provide some form of read back. However the controller does not always require a complete read back.
- **Hold Short**: pilots almost always correctly read back both the hold short instruction and position and controllers may always require a correct and complete read back to this command

Not all discrepancies in a read back warrant an alert.

A good design will consider the risk associated with the read back error.

MITRE

Speech Recognition Evaluation



Engine: Sphinx4

Digitized Audio Signal

Acoustic Model

Three types of acoustic models were evaluated:

- A base English acoustic model released by CMUSphinx
- Adaptations of the English acoustic model, enhanced using controller and pilot recordings and transcriptions
- Custom acoustic models created using only controller and pilot recordings

Basic Speech Sounds

… r ʌ n w e ɪ w ʌ n …

Language Model

Two statistical language models were created

- One trained using transcriptions of local and ground controller radio transmissions
- One trained using transcriptions of pilot radio transmissions heard at the pilot and ground controller positions

Speech Hypothesis

… **three hold short of runway one niner right** …

Command Concept

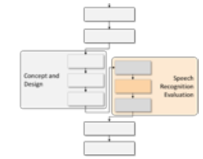
… | Hold Short | 19R | …

A semantic parse algorithm was used to extract command concepts and aircraft identifiers that were robust to word variations

MITRE

Speech Recognition Evaluation

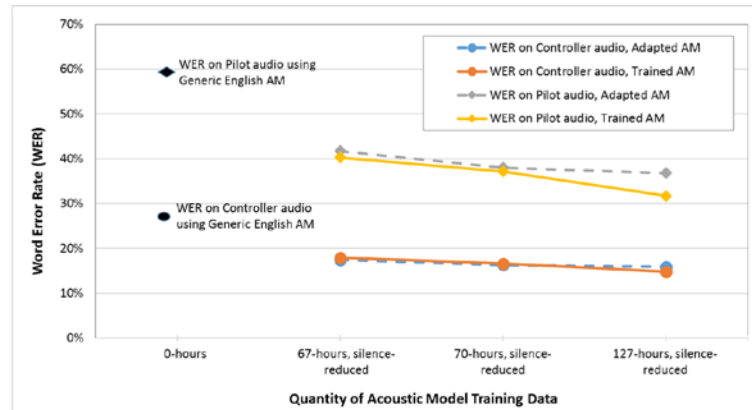
Model Comparison



Word Error Rate (WER) was used as a rough performance measure to compare different speech recognition tuning methods and configurations

$$Word\ Error\ Rate = omissions + additions + substitutions$$

- ✓ **Custom acoustic model performed best**
- ✓ **Bigger advantage on pilot speech as size of training set increases (see right)**



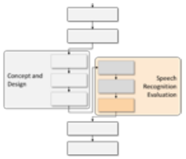
- *Test data consisted of 10 hours of audio recordings from JFK containing clearance and read back exchanges*
  - *3,689 pilot transmissions*
  - *2,970 controller transmissions*
- *Training data consisted of 127 hours of transcribed, silence-reduced audio from the air traffic control tower domain*
  - *Several facilities (including JFK)*
  - *Both controller and pilot speech*

But concept recognition is a better predictor of application performance

Command Concept	Nominal Command Phrase	List of Parameter Alternatives
CTL	“cleared to land”	Runway
LUAW	“line up and wait”	Runway
CFT	“cleared for takeoff”	Runway
Hold Short	“hold short”	Runway or Taxiway
Cross	“cross”	Runway or Taxiway
Turn	“turn left” or “turn right”	Runway or Taxiway
Taxi	“taxi”	Runway(s) or Taxiway(s),
Continue	“continue”	Runway(s) or Taxiway(s)

Speech Recognition Evaluation

Concept Recognition Accuracy

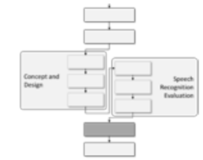


- Better concept recognition of controller speech than of pilot speech
- ACID recognition in this test is poor; previously, we have demonstrated better results with restrictive context-informed grammar.
- The addition of dialogue context across transmissions could improve clearance to read back matching, despite low ACID recognition

Controller Transmissions	ACID	LUAW	CFT	CTL	Hold Short	Cross	Turn Right	Turn Left	Taxi	Continue
% Recognized Correctly	79%	94%	98%	91%	96%	96%	98%	97%	93%	97%
% Partially Recognized	NA	5%	1%	1%	3%	3%	2%	1%	2%	3%
% Not Recognized	21%	1%	2%	7%	1%	1%	0%	2%	5%	0%

Pilot Transmissions	ACID	LUAW	CFT	CTL	Hold Short	Cross	Turn Right	Turn Left	Taxi
% Recognized Correctly	63%	86%	80%	84%	88%	92%	87%	84%	80%
% Partially Recognized	NA	7%	8%	3%	8%	4%	4%	5%	12%
% Not Recognized	37%	8%	12%	13%	4%	5%	9%	11%	7%

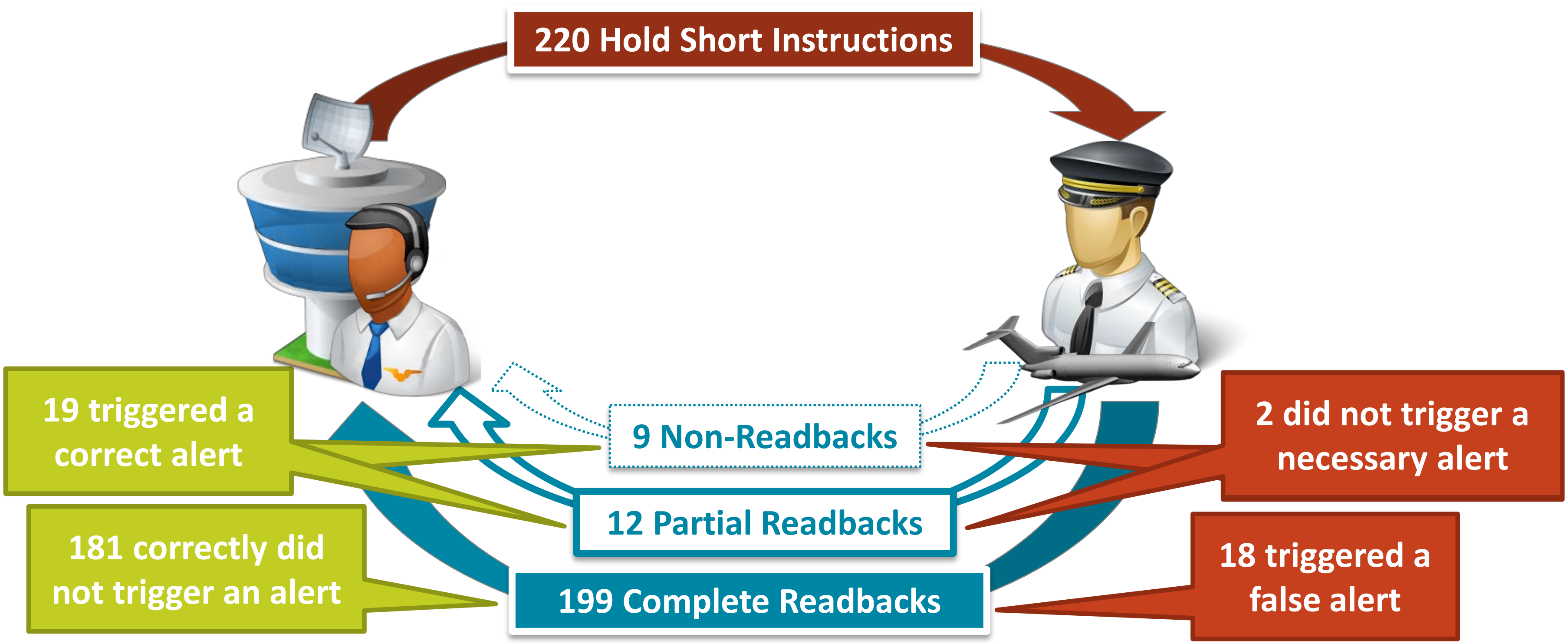
But application performance driven by concept recognition in sequence, not aggregate (see next slide)



Case Study: Hold Short

A look at potential system performance with current speech recognition performance on the Hold Short instruction data set

- Alert logic only considered controller clearance and pilot readback pairs
- Required readback of both command and runway



MITRE

Key Takeaways

- **Not all read back discrepancies warrant an alert**
- **Key to application design is appropriate consideration of what warrants an alert**

- **Using real ATC voice data enables better speech recognition accuracy**
- **Recognizing pilot audio is particularly challenging**

- **Assessing the application requires both speech recognition accuracy <u>and</u> a set of alert rules, and the two must be set in consideration of each other**

Fiscal Year:	2017
Outcome Number:	8
PBWP Reference:	8-1.1