

Grammar Based Identification Of Speaker Role For Improving ATCO And Pilot ASR

Amrutha Prasad, Juan Zuluaga-Gomez, Petr Motlicek, Oliver
Ohneiser, Hartmut Helmke, Saeed Sarfjoo, Iuliia Nigmatulina



- Amrutha Prasad
 - Masters in Artificial Intelligence (2019-2020)
 - PhD student at Brno University of Technology
 - Research Assistant at Idiap Research Institute



Motivation

- Previous research focused only on ATCO automatic speech recognition (ASR)
- Current focus is to develop both ATCO and pilot ASR
- Challenges for pilot ASR:
 - Acoustic and grammatical conditions are different
 - Less data available
 - Speaker role (ATCO or pilot) not available in data

Speaker role classification

- Use grammar to classify an utterance as ATCO or pilot.
- The words such as "identified", "approved", "wind" would most probably only be spoken by an ATCO.
- The words "wilco", "maintaining", "we", "our" would probably be spoken only by a pilot.
- Current list: 25 words for ATCO and 9 words for pilot.

Speaker Role Classification Results

- Tested on manually speaker segmented and transcribed data
 - NATS for London Approach: 1060 ATCO utterances and 1280 pilot utterances.
 - ISAVIA for Icelandic en-route: 775 ATCO utterances and 887 pilot utterances.

NATS

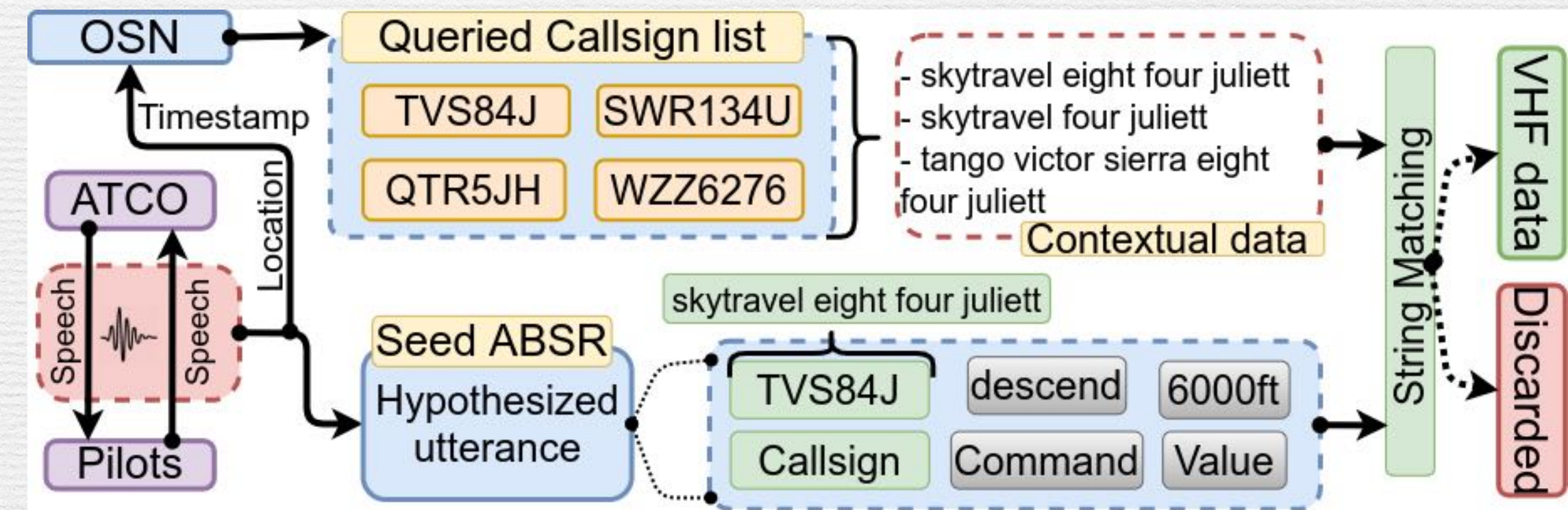
Predicted Class	Actual	
	ATCO	Pilot
ATCO	856 81%	188 15%
Pilot	204 19%	1092 85%

Isavia

Predicted Class	Actual	
	ATCO	Pilot
ATCO	660 85%	179 20%
Pilot	115 15%	708 80%

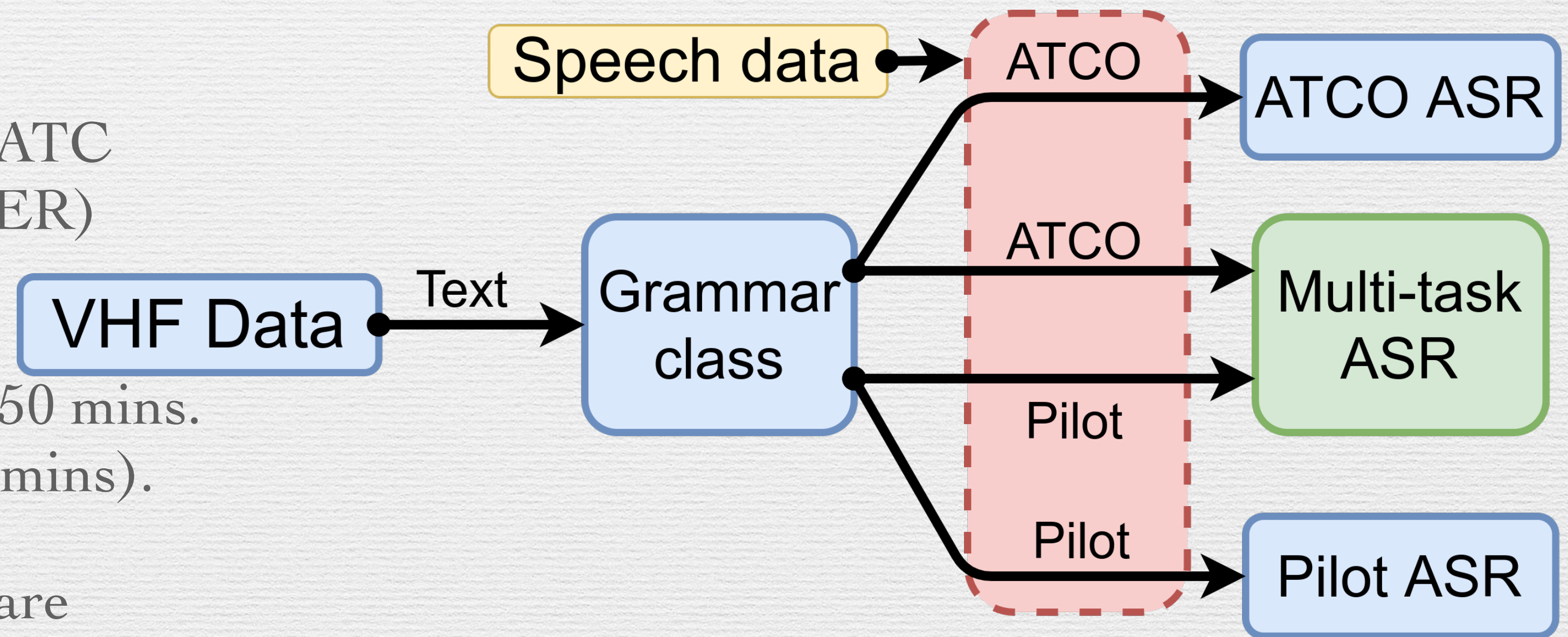
Datasets

Data	Duration (h)	Description
VHF	377-> 66	<ul style="list-style-type: none"> - Open source (ex: Live ATC) - VHF receiver
ATC related	140	<ul style="list-style-type: none"> - Hiwire: noisy and non-native cockpit communication - Atcosim: ATCO speech - French accent ATC - MALORCA: ATCO dataset - Military ATC - LDC databases



Experiments

- AMs are trained with the state-of-the-art LF-MMI training framework. systems.
- The performance is evaluated on LiveATC test set with the Word Error Rate (WER) metric.
- The total duration of the test set is 1h 50 mins. ATCO set (52 mins) and Pilot set (58 mins).
- In each group of experiments, results are given for i) AM trained for each task separately, ii) AM trained by combining all data and iii) AM trained with multitask learning.



Results

Model	ATCO WER (%)	Pilot WER (%)
VHF ATCO	43.2	51.6
VHF Pilot	40.3	45
Combined	46	50
Multitask	38.2	44

Model	ATCO WER (%)	Pilot WER (%)
ATCO	30.3	43.2
Pilot	32.8	40.3
Combined	31.2	41.3
Multitask	31.9	41.3

Conclusion

- A simple grammar based approach to identify speaker role, accuracy of approx. 85%
- Train acoustic models either by speaker role or in a multitask fashion.
- Multitask training approach outperforms other training methods when limited training data is available.
- Training AMs separately provides better ASR performance when sufficient data is available.
- Relative improvements of 3.2% for the ATCO set and 1.9% for the pilot set were obtained.

Thank you