Improving callsign recognition with air-surveillance data in air-traffic communication

Interspeech, BUT, Brno, 30.08.21
Satellite Workshop – Automatic Speech Recognition in Air Traffic Management (ASR-ATM)

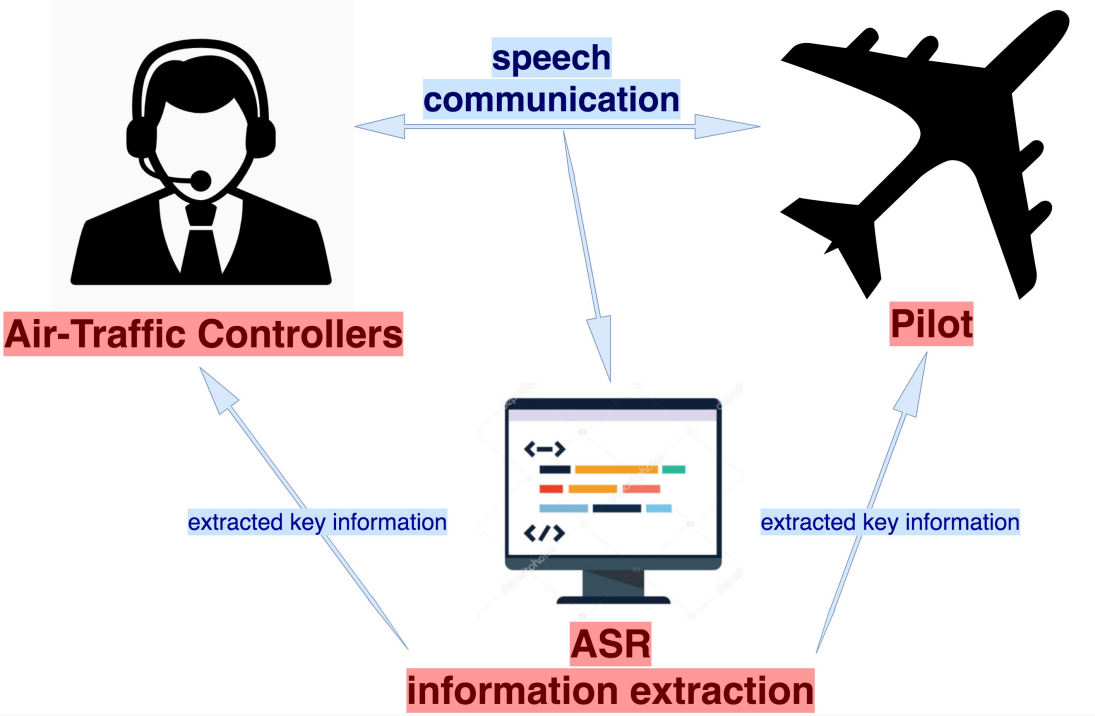
Iuliia Nigmatulina, Rudolf A. Braun, Juan Zuluaga-Gomez, Petr Motlicek

Curriculum Vitae

Iuliia Nigmatulina

- PhD student at the Idiap Research Institute (since January 2021)
- Master degree in Phonetics (2013)
- Master degree in Computational Linguistics and Speech Technologies (2020)
- currently, working for ASR and ATC projects

1.1. ASR for Air-Traffic communication



Automatic speech recognition assistance can help.
But **high accuracy** of the key information is crucial.

1.2. Callsigns

Callsign is a unique identifier for aircraft, of which the first part is an abbreviation of airline name and the last part is a flight number of digits and letters, where letters are encoded with special words.

- **SWR2689** - swiss two six eight nine
- **RYR1RK** - ryanair one **romeo kilo**
- **RYR1SG** - ryanair one **sierra golf**

1.3. Radar: contextual information

2016-08-13_10-09-06-12 BEL85A CSA3CT CSA94D EZY7905 FDB779 GMI6452 HOP4412

Goal:

to increase the probability of recognising those **callsigns** which are present in the air space at the moment of utterance by **dynamically** introducing **contextual information (radar)**.

1.4. Methods

Boosting callsigns = boosting **n-grams** by dynamically modifying their weights in the weighted finite-state transducer (WFST) with the **FST composition**.

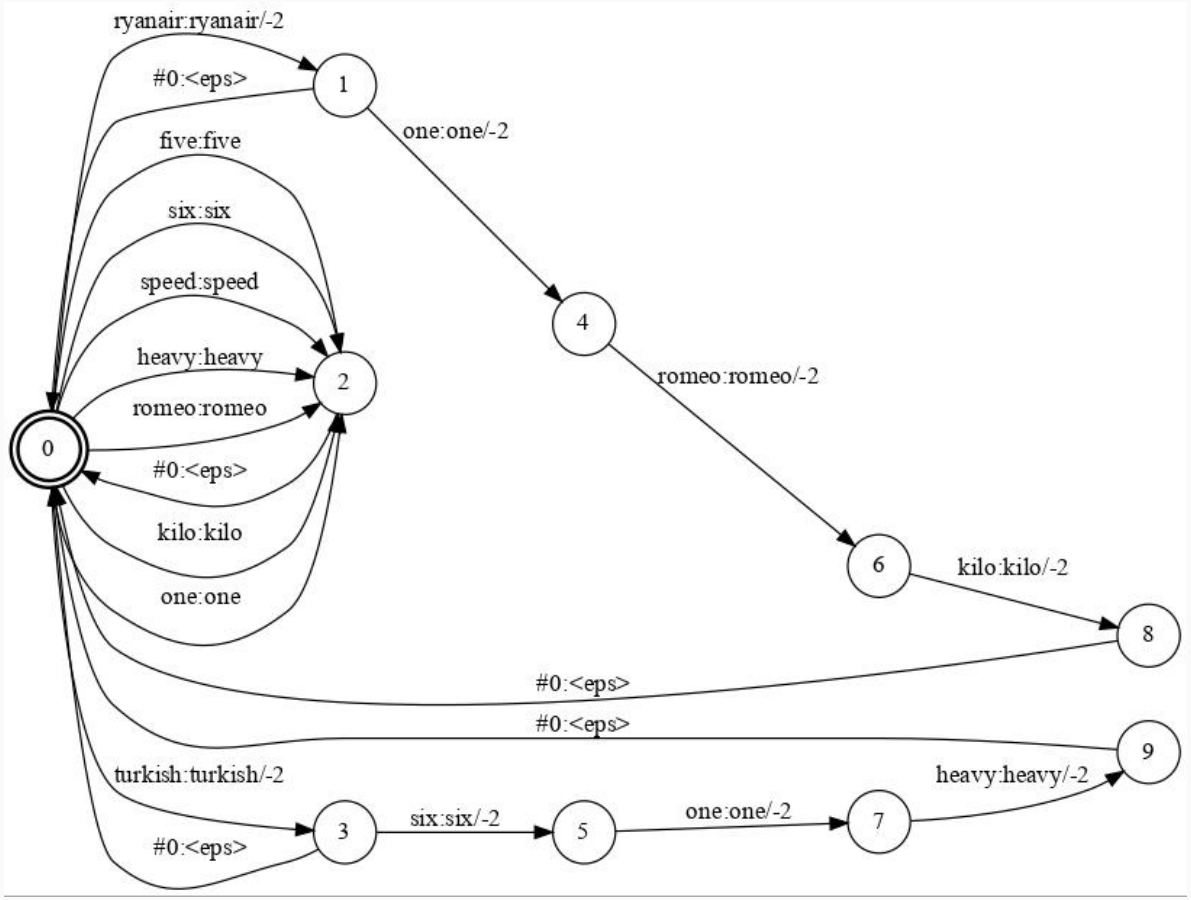
Two approaches:

- ***Lattice rescoring*** to boost specific callsigns (composition after decoding).
- ***Grammar modification*** with boosting n-grams weights in G.fst (composition before decoding).

In both approaches, composition is done **per utterance**.

1.5. Lattice rescoring

$$final\_decoding\_FST = lattice\_FST \circ biased\_FST \tag{1}$$



Biased FST with callsigns: **'ryanair one romeo kilo'** and **'turkish six one heavy'**.

1.6. G-boosting

G_biased is a baseline G.fst (language model) with weights adjusted to boost callsign n-grams and new callsign n-grams added.

$$HCLG\_biased = HCL \circ G\_biased \tag{2}$$

HCLG_biased is composed ***on-the-fly*** per utterance ***during the decoding***.

1.7. Data

Test set	Num of utterances	Utterances with callsigns (%)	Callsigns per utterance (median)	Minutes
LiveATC_mix[1]	610	95%	28	40
Malorca[2] Prague	872	90%	5	82
Malorca Vienna	915	96%	19	65

[1]LiveATC.net is primarily a streaming audio network consisting of local receivers tuned to aircraft communications around the world:https://www.liveatc.net/.

[2]The Horizon 2020 SESAR project MALORCA (Machine Learn-ing of Speech Recognition Models for Controller Assistance) ispartly funded by SESAR Joint Undertaking (Grant Number 698824):https://www.malorca-project.de/wp/.

1.8. Setup

- Kaldi framework [3];
- CNN-TDNNF trained on approximately 1200 hours (after noise augmentation and perturbation);
- lexicon: 28410 words;
- 3-gram LM;
- *baseline*: the model without applying any boosting mechanisms.

Evaluation:

- Word Error Rate (WER) on a full utterance
- callsign WER
- callsign accuracy

[3] Povey, D., A.Ghoshal, G.Boulianne, L.Burget, O.Glembek, N.Goel, M.Hannemann, P.Motlicek, Y.Qian, P.Schwarz, et al. (2011). "The Kaldi speech recognition toolkit". In: IEEE workshop on automatic speech recognition and understanding. CONF. IEEE Signal Processing Society.

1.9. Results

Table 1: Results of the boosting experiments (WER — word error rate; CWER — Callsign WER; Acc — accuracy of callsign recognition)

	LiveATC_mix			Malorca Prague			Malorca Vienna		
Model	**WER**	**CWER**	**Acc**	**WER**	**CWER**	**Acc**	**WER**	**CWER**	**Acc**
baseline	30.7	29.2	50.5	3.1	2.2	94.2	9.2	6.6	84.6
lattice rescor.	29.5	23.9	60.8	**3.0**	**1.0**	97.	8.3	3.1	93.8
G boosting	28.1	19.5	66.2	3.1	1.7	95.7	8.5	3.6	91.9
G+lattice rescor.	**27.2**	**16.0**	**71.3**	3.1	**1.0**	**97.4**	**8.2**	**1.7**	**96.3**
ground truth	26.3	12.2	79.8	2.8	0.8	97.7	8.1	1.4	97.6

1.10. Improvements

Table 2: Examples of improved callsign recognition (red — wrong; blue — correct)

Callsign	System	Callsign expanded
STK19L	Baseline	hello sovar one nine lima
	Boosted	stobart two one nine lima
RYR4TM	Baseline	ryanair four bye bye
	Boosted	ryanair four tango mike
AFR6735	Baseline	one six zero three five
	Boosted	airfrans six seven three five

1.11. Discussion

G-boosting:

- ☺ biased weights before building lattices.
- ☹ more memory consuming: keeping at least two G.fst.

Lattice rescoring:

- ☹ rescoring is done after lattices are built.
- ☺ less memory consuming; easy to implement in the **online recognition** with no significant latency (tested with biasing FST including 30 n-grams).

Conclusion

- The best results — with the **combination of both methods**.
- The improvement is noticeable in all test sets: from 45.2 to 74.2% of relative improvement in callsign WER depending on the test set.
- Lattice rescoring can be used in **online recognition**.