

CZAI ASR workshop

Igor Szoke, Veronika Nevařilová (9.9.2026)



- use cases a co znamenají pro volbu ASR.
- Trénovacích dat (identifikace, akvizice), jejich případná anotace a čištění.
- Příprava trénovacích a evaluačních setů
- Metrik vyhodnocení.
- Nástrojům, modelům a jejich adaptaci na cílovou doménu.
- Jak model naadaptovat
- následně jak provozovat.





DATE 7/11	POINT ↑	POINT TIME	REG OK UHE	OPR	POB 7	RMSK
TWY RWY B 200	FL	FL	TYPE SR22	WTC L	DEP KM	RFL FUEL SQ
DATE 7/11	POINT ↓	POINT TIME	REG OK HLA	OPR	POB 4	RMSK W, VEL, U4
TWY RWY B 200	FL	FL	TYPE C142	WTC L	DEP KM	RFL FUEL 3.8
DATE 7/11	POINT ↑	POINT TIME	REG OK NUF	OPR	POB 2	RMSK E
TWY RWY B 200	FL	FL	TYPE SR22	WTC L	DEP KM	RFL FUEL 2.9
DATE 7/11	POINT ↑	POINT TIME	REG OK EURON	OPR	POB 7	RMSK VESELF
TWY RWY B 200	FL	FL	TYPE P280	WTC L	DEP KM	RFL FUEL 4.8
DATE 7/11	POINT ↓	POINT TIME	REG OK VAV 99	OPR	POB	RMSK
TWY RWY B 200	FL	FL	TYPE UAC	WTC L	DEP KM	RFL FUEL 4.4
DATE 7/11	POINT ↑	POINT TIME	REG OK QUU06	OPR	POB 7	RMSK AD
TWY RWY B 200	FL	FL	TYPE NG5	WTC L	DEP KM	RFL FUEL 15.8
DATE 7/11	POINT ↓	POINT TIME	REG OK ACT	OPR	POB 5	RMSK
TWY RWY B 200	FL	FL	TYPE P46T	WTC L	DEP PR	RFL FUEL 2.4
DATE 7/11	POINT ↑	POINT TIME	REG OAM CSD	OPR	POB 2	RMSK
TWY RWY B 200	FL	FL	TYPE W73	WTC L	DEP PR	RFL FUEL 7.8
DATE 7/11	POINT ↓	POINT TIME	REG OK 0105	OPR	POB 2	RMSK
TWY RWY B 200	FL	FL	TYPE VIVAT	WTC L	DEP KM	RFL FUEL 4.4

Kunovice Věž Oscar Kilo Zulu Uniform Romeo Jedna Čtyři dobré ráno

Oscar Zulu Romeo Jedna Čtyři Kunovice Věž

Kunovice Věž je to ultralehký sportcruiser ze Slušovic k vám do Kunovic jedna osoba na palubě QNH jedna nula jedna tři na dva tisíce tři sta feetů a žádám o povolení ke vstupu do CTR a na přistání

Oscar Zulu Romeo Jedna Čtyři vstup do CTR povolen QNH jedna nula jedna tři pokračujte na dlouhé finále dráhy dva nula právě vítr klid

pokračuji na dlouhé finále dva nula právě QNH jedna nula jedna tři Oscar Kilo Zulu Uniform Romeo Jedna Čtyři



DATE 3/9		POINT	POINT	CALLSIGN OKZUR14		V 	OPR	POB 1
		TIME	TIME	REG	SQUAWK		RFL	FUEL
TWY	RWY 20R	FL A23	FL	TYPE ULAC	WTC	DEP LKSLUS	TIME	DEST LKKU

+ DEP

+ ARR

+ LOC

+ OVR

Hlavní

Archiv

Koš

TWR

Obnovit

Online

Aktuální lety

4 lety

DATE	↑↓	POINT	POINT	CALLSIGN		V	OPR		POB		Buchlov	
1/1		TIME	TIME	REG			SQUAWK		RFL			FUEL
TWY	B	RWY	20C	FL	FL	TYPE	WTC	DEP	TIME	DEST		TIME
						C172	L	LKKU	1057	LKKU		

DATE	*	POINT	POINT	CALLSIGN		Z	OPR		POB		Remarks...	
1/1		TIME	TIME	REG			SQUAWK		RFL			FUEL
TWY		RWY		FL	FL	TYPE	WTC	DEP	TIME	DEST		TIME
						ULAC	L	LKTB		LZIB		

DATE	↑↓	POINT	POINT	CALLSIGN		V	OPR		POB		Remarks...	
1/1		TIME	TIME	REG			SQUAWK		RFL			FUEL
TWY	B	RWY	20R	FL	FL	TYPE	WTC	DEP	TIME	DEST		TIME
						ULAC	L	LKKU	1042	LKKU		1049

DATE	↑	POINT	POINT	CALLSIGN		V	OPR		POB		Remarks...	
12/1		TIME	TIME	REG			SQUAWK		RFL			FUEL
TWY		RWY	20R	FL	FL	TYPE	WTC	DEP	TIME	DEST		TIME
						L23	L	LKKU	1042	LKKU		1051

Pojďme vzít dostupný model.

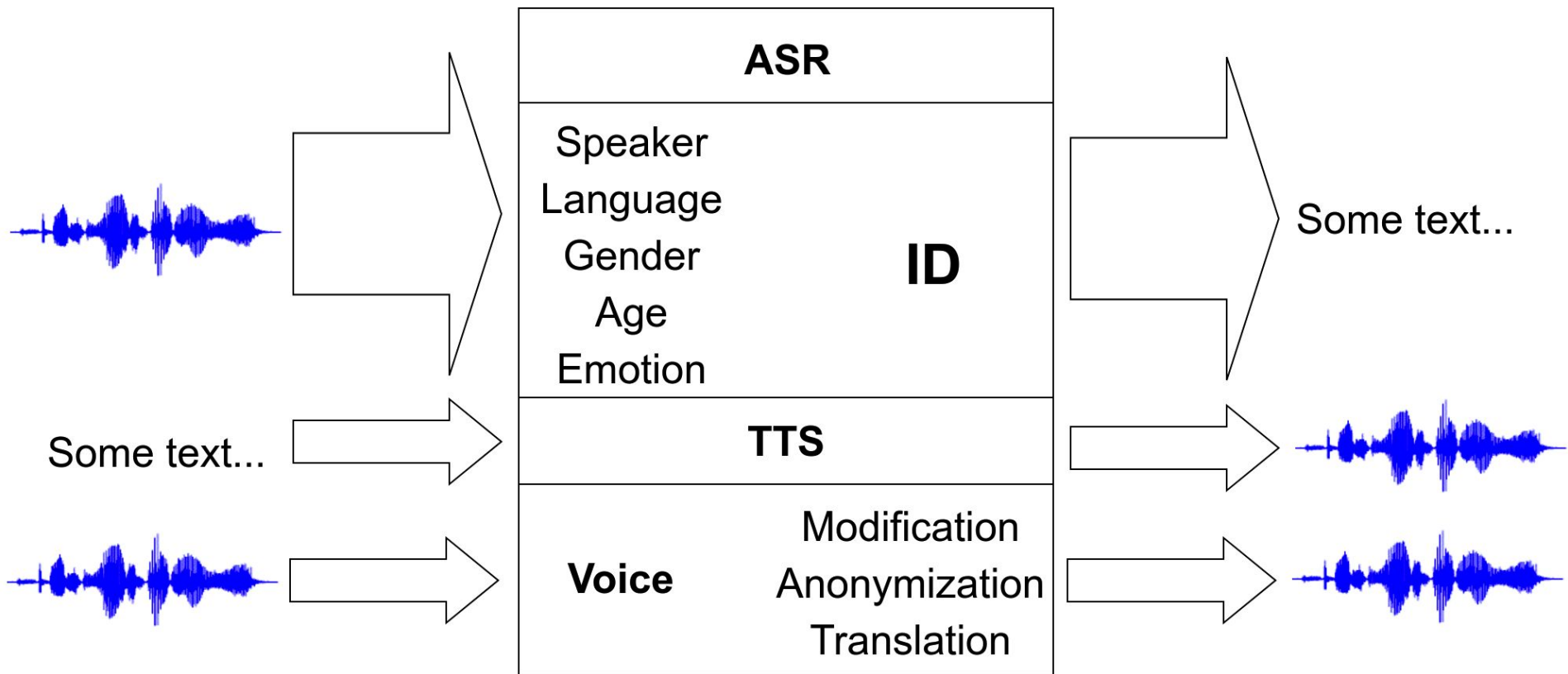


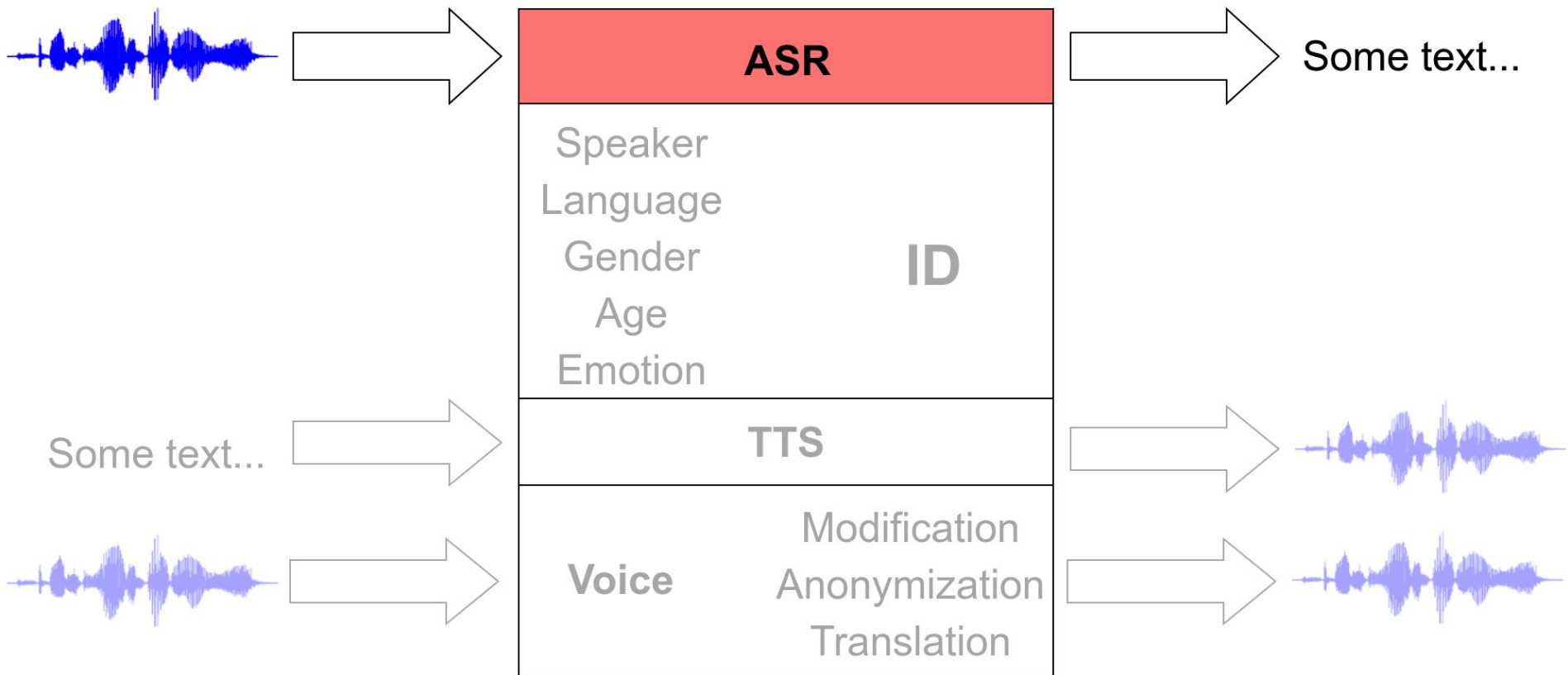
Pojďme vzít dostupný model.

 **Whisper Large** → ATCO: RY4Z, altitude 4000ft, QNH 1008
Pilot: **ATC talking to the radio**

 **Whisper Large** → ATCO: **ATC talking to the radio**

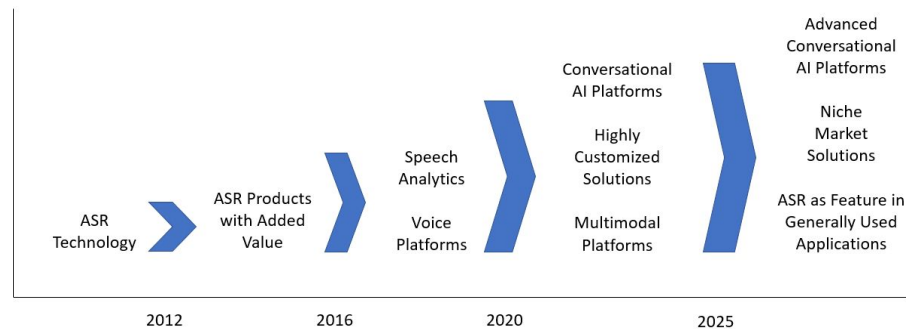
OK ... tudy cesta nevede





Proč chci ASR?

- Chci přepis?
 - Titulky, diktování
- Chci vytěžovat informace?
 - Sumarizace, analytika
- Chci komunikovat?
 - Plechová huba, ovládání něčeho

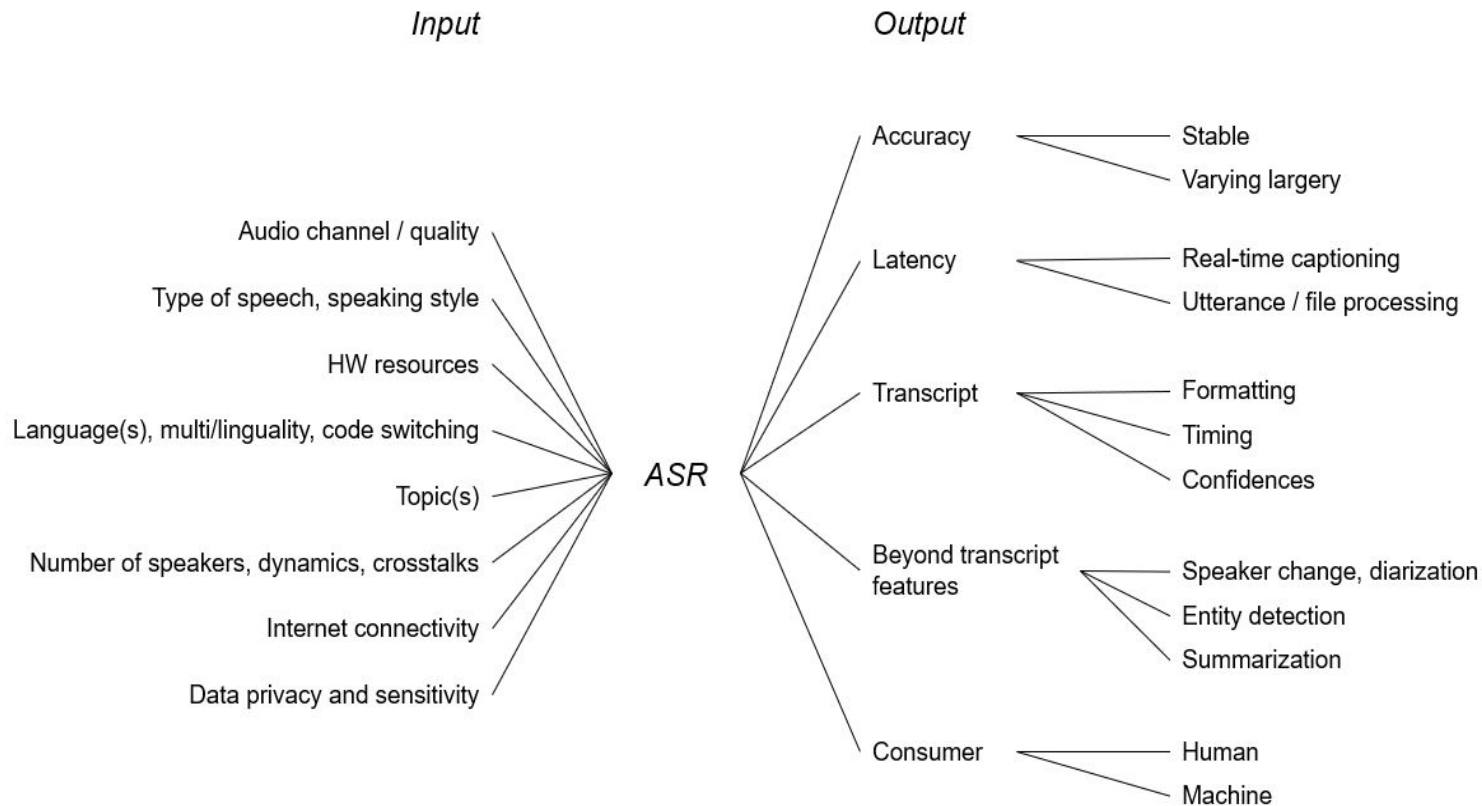


Market

- Accuracy
- Data Privacy & Ethics
- Regulation
- Competition
- Integration
- Cost

Technical

- Accuracy / Word error rate (WER)
- Training data, robustness
- Latency
- CPU/RAM footprint and deployment
- Storage of the user data
- Traceability, explainability and personal data protection
- Adaptability
- Output format
- Used technology / model complexity / NN architecture
- Implementation environment



Knowledge

Architecture

Tools

The ASR

Model

Data

Hardware

Time

Knowledge

Architecture

Tools

The ASR

Model

Data

Hardware

Time

Data

Mám data?

Mám cílová data?

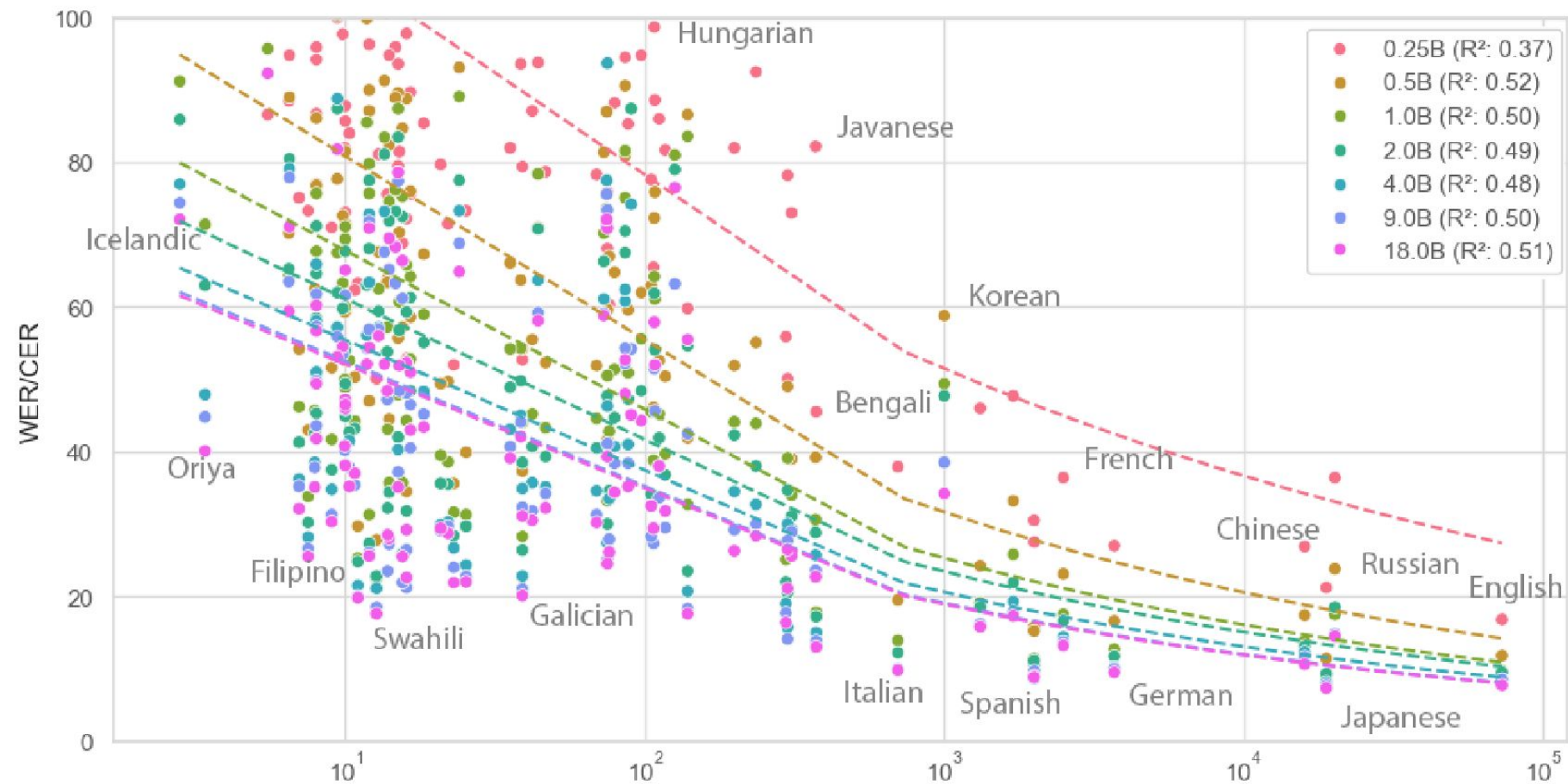
Pokud ne, tak musíte nějaká data vyrobit, stáhnout (nebo koupit):

- stažení existujících
- anotace interních existujících dat - tento proces může vzít 2 - 20 RT.
- výroba syntetických dat
 - “zapravení” existujících
 - výroba nových

- ELRA, ELRA Language Resources Association, <http://www.elra.info/en/>
- LDC, Linguistic Data Consortium, <https://www.ldc.upenn.edu/>
- CLARIN, the research infrastructure for language as social and cultural data, <https://www.clarin.eu/>
- SpeechOcean, <https://en.speechocean.com/>
- Data Tang, <https://www.datatang.ai/>
- ZENODO, <https://zenodo.org/>
- Appen, <https://appen.com/>
- META-SHARE, <http://www.meta-share.org/> and now the ELG platform, <https://live.european-language-grid.eu/>
- OLAC (a catalog of catalogs), the Open Language Archives Community, <http://www.language-archives.org/>
- but also Google Data Search, Huggingface etc.

A pak samozřejmě data scraping...

ALL THESE LICENCES ARE PERPETUAL, WORLDWIDE, ALLOWING REDISTRIBUTION OF THE ORIGINAL INSIDE & OUTSIDE THE METASHARE NETWORK													
Licence name	Licence Attributes	Should the Original be redistributed within the META SHARE Network only?		Is Attribution required?		What purposes can the Original Resource be used for?		Are Derivatives allowed?		Do derivative have to be shared under the same terms?		Is there any Fee for this Licence?	
		Yes	No	Yes	No	Commercial	NonCommercial	Yes	No	Yes	No	Yes	No
CC-ZERO			✓		✓	✓		✓			✓		✓
CC BY	BY		✓	✓		✓		✓			✓		✓
CC BY SA	BY SA		✓	✓		✓		✓		✓			✓
CC BY ND	BY ND		✓	✓		✓			✓	N/A	N/A		✓
CC BY NC SA	BY NC SA		✓	✓			✓	✓		✓			✓
CC BY NC	BY NC		✓	✓			✓	✓			✓		✓
CC BY NC ND	BY NC ND		✓	✓			✓		✓	N/A	N/A		✓

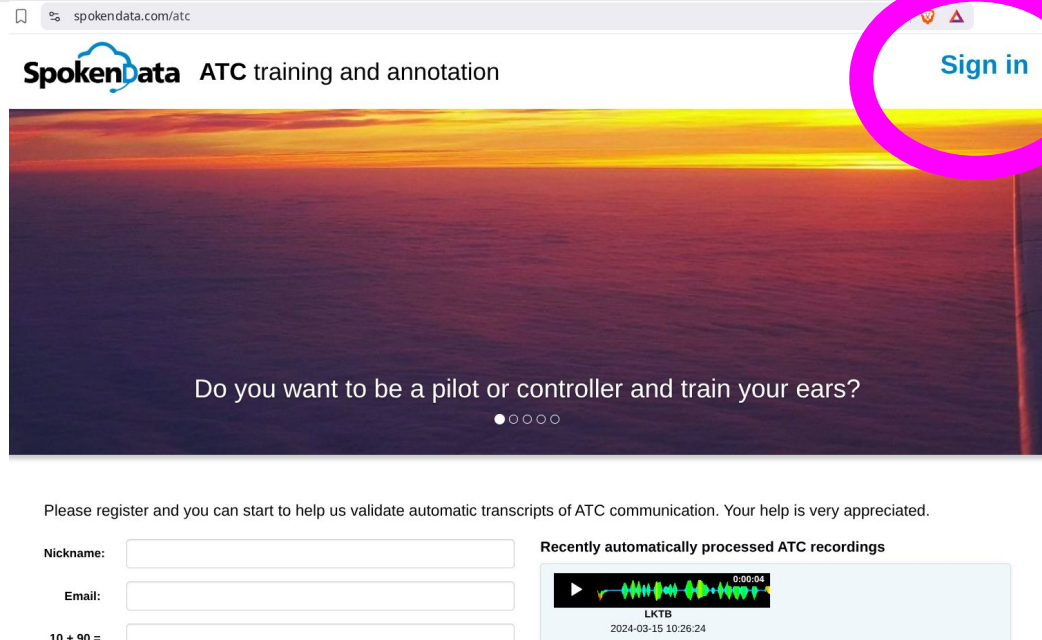


https://spokendata.com/atc

usr: tomas@szoke.cz

pwd: heslo4CZAI

Zkuste si něco oanotovat
(sluchátka doporučena)



spokendata.com/atc

SpokenData ATC training and annotation

Do you want to be a pilot or controller and train your ears?

Please register and you can start to help us validate automatic transcripts of ATC communication. Your help is very appreciated.

Nickname:

Email:

10 + 90 =

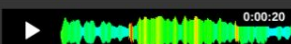
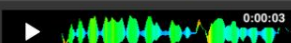
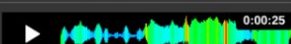
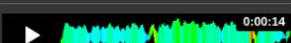
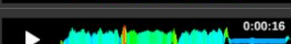
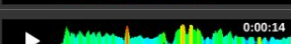
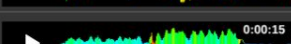
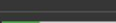
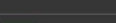
Recently automatically processed ATC recordings

 0:00:04

LKT8

2024-03-15 10:26:24

My recordings

LKTB BRNO 2020-08-19 13:31:33					
LKTB BRNO 2020-08-18 15:04:41					
LKTB BRNO 2020-08-15 19:36:16					
LKTB BRNO 2020-08-15 10:52:08					
LKPR LKPR_Radar_119_000MHz - 2023-12-02 02:28:01 UTC					
LKPR LKPR_Tower_134_560MHz - 2023-12-01 19:36:05 UTC Czech					
LSZB LSZB_BERN_Approach_127_3MHz - 2023-09-01 13:28:01 UTC					
LSZB LSZB_BERN_Tower_121_0MHz - 2023-09-01 13:28:43 UTC					
LKPR LKPR_Radar_120_520MHz - 2023-08-31 08:42:34 UTC Czech					
LSZH LSZH_ZURICH_ApronN_121_85MHz - 2022-10-22 14:20:44 UTC					
LKPR LKPR_Tower_134_560MHz - 2022-09-04 08:13:51 UTC					
LSZB LSZB_BERN_Approach_127_3MHz - 2022-05-20 14:27:36 UTC					
LSZB LSZB_BERN_Tower_121_0MHz - 2022-05-12 09:21:22 UTC					

🔍

🔍

Playback speed: Normal

🔄

⏮

▶

⏭

00:11.5 / 00:15.1

0

5

10

Changes

🔄

🔄

📄 Export

💾 Save

Since last save: 00:00:58

Job control

✓ Close as REFUSED

✓ Close as DONE

Text labels

Section labels:

1 CALLSIGN

2 COMMAND

3 VALUE

4 UNNAMED

5 ANONYMIZE

6 NONENGLISH

Transcription tags:

7 UNKNOWN

8 HESITATION

9 NOISE

10 SPEAKER NOISE

11 DOUBLE-PRESS PTT

12 CROSSTALK

▶ 00:00.3
00:05.8

Departure good day it's Pad Aviation Three One Kilo passing
three thousand one hundred climbing five thousand
on departure route

PVD31K

▶ 00:05.9
00:11.5

Pad Aviation Three One Kilo Bern Departure hello climb to
flight level one hundred now identified

Departure

▶ 00:11.7
00:15.0

climb level one hundred(hundert) roger
Pad Aviation Three One Kilo

PVD31K

Job details

Name: LSZB
LSZB_BERN_Approach_127_3MHz
- 2022-05-20 14:27:36 UTC

Description:
filename:
"LSZB_BERN_Approach_127_3MHz_

Display attached resources:
Waypoints ☐
Sorted Waypoint-Callsign pairs ☐

Manuals and others: Attached links

Speaker labels

A Departure

B PVD31K

+ Add label

23

Co znamená “Dobré ASR”?

- Objektivní metrika/metriky.
- Snadno měřitelné (integrace do procesu trénování).
- Robustnost

Obvykle se měří (optimalizuje) na WER a CER. (koreluje s downstream komponenty).

departure good day it's pad aviation three one kilo passing three thousand one hundred climbing five thousand on departure route

VS.

Departure, good day. It's PVD31K passing 3100 climbing 5000 on departure route.

El. kniha letů

- WER nemusí být správná/směrodatná metrika
- Nutná správnost především klíčových slov

Oscar Kilo Alpha Bravo Charlie Cessna one seven two at entry point Sierra altitude two thousand feet and we will continue to right baseleg for runway zero two center

Train / Dev / Eval subsets

- Eval - absolutně neviděná cílová data - pár hodin, ideální jsou i subsety
 - pokud mám podivné “drobky”, malé řečníky, atd.
- Dev - netrénuje se, ale optimalizuje (obvykle část Train)
 - trénovací hyper parametry
- Train - vše co mám

Když mám málo dat?

- Leave-One-(Speaker)-Out Cross-Validation (rotace subsetů a průměr WER)

Architecture / Tools / Model

Name (main driver)	Community	Neural network backend	Streaming ASR supported	Hybrid / E2E	License	Reference
Kaldi (JHU, D. Povey)	Large but decreasing	Kaldi	YES	Hybrid	Apache 2.0	https://kaldi-asr.org/
ESPnet (CMU, S. Watanabe)	Large	PyTorch	YES (partially)	Hybrid, E2E	Apache 2.0	https://github.com/espnet/espnet
k2-fsa (Xiaomi, D. Povey)	Small but increasing	PyTorch	YES	E2E	Apache 2.0	https://github.com/k2-fsa/k2 (training) https://github.com/k2-fsa/sherpa (runtime)
SpeechBrain (MILA, M. Ravanelli)	Medium	PyTorch	NO	E2E	Apache 2.0	https://github.com/speechbrain/speechbrain
NeMo (Nvidia)	Large	PyTorch Lightning	YES (partially)	E2E	Apache 2.0	https://github.com/NVIDIA/NeMo
Athena (Beike AAC, Beijing)	Small and decreasing	TensorFlow	YES	E2E	Apache 2.0	https://github.com/athena-team/athena
Espresso (JHU, Yiming Wang)	Large but decreasing	PyTorch	NO	E2E	MIT License	https://github.com/freewym/espresso
WeNet (Horizon Robotics, B. Zhang)	Medium and increasing	PyTorch	YES	E2E	Apache 2.0	https://github.com/wenet-e2e/wenet

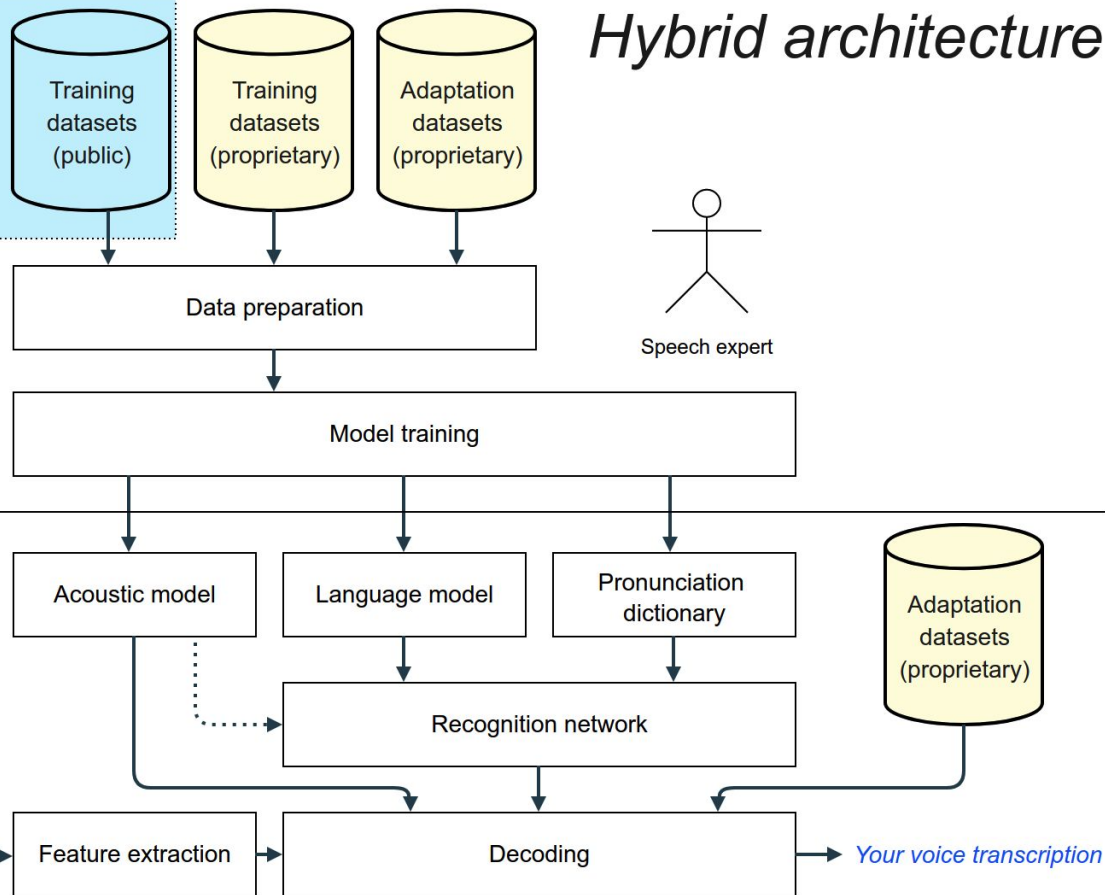


Jaké máme omezení / požadavky na e-stripy?

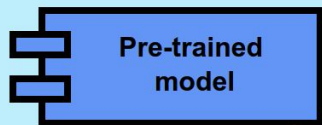
- Training
 - několik GPUs
 - Python
 - podpora
- Runtime (inference)
 - batch
 - čeština a angličtina
 - +- 1RT je OK
 - CPU
- Pretrained models
 - Ano, potřebujeme

Publicly available

Hybrid architecture



Publicly available



*E2E architecture
pre-trained*

Training

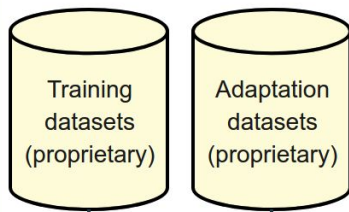
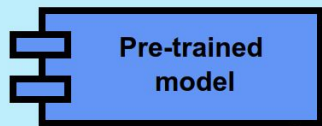
Runtime



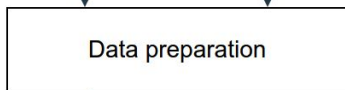
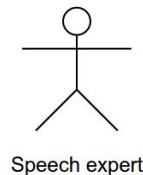
End2End recognition model and decoding

Your voice transcription

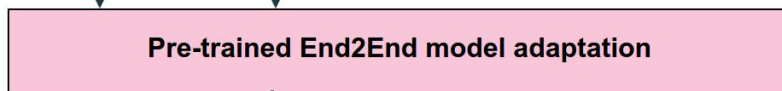
Publicly available



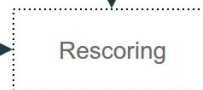
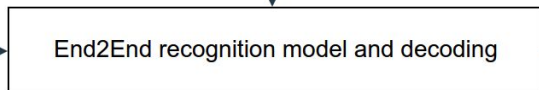
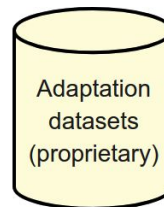
*E2E architecture
pre-trained*



Training



Runtime



Your voice transcription

End-to-end ASR models



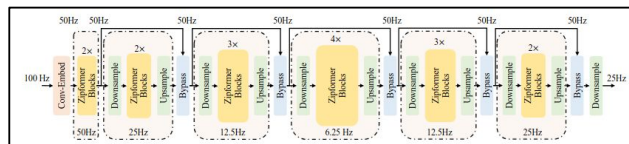


Figure 1: Overall architecture of Zipformer.

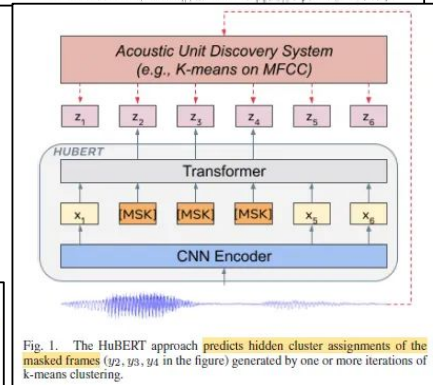
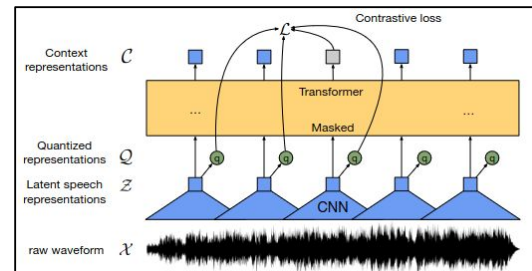
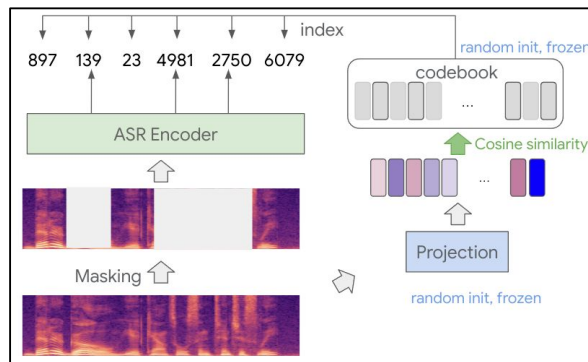
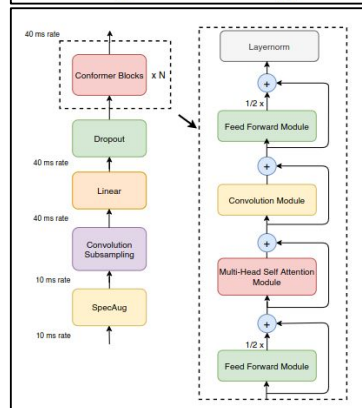
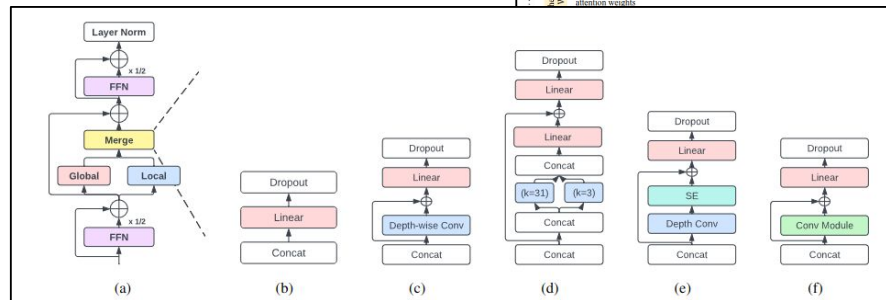
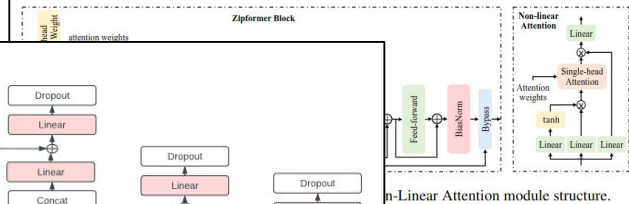
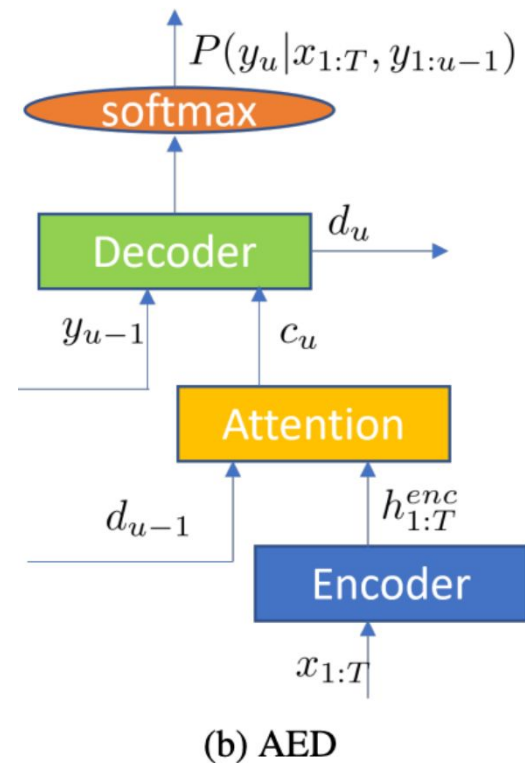
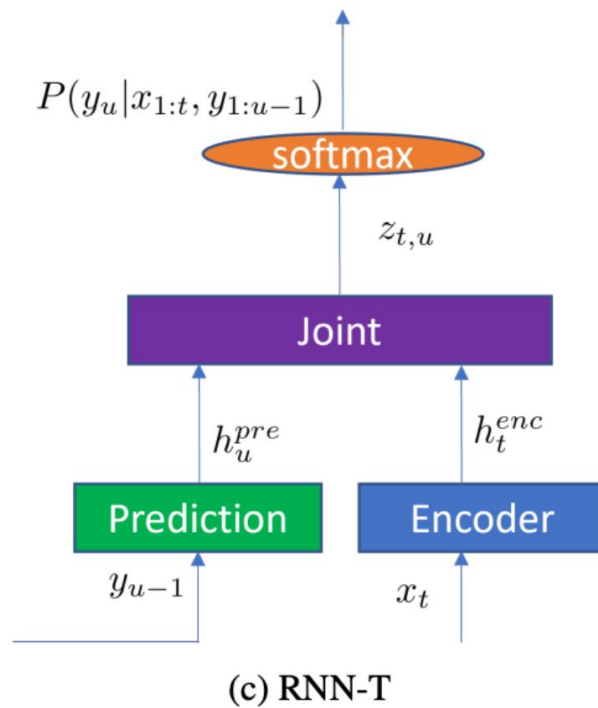
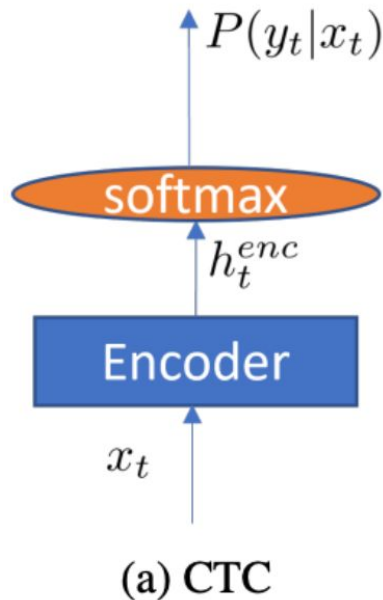
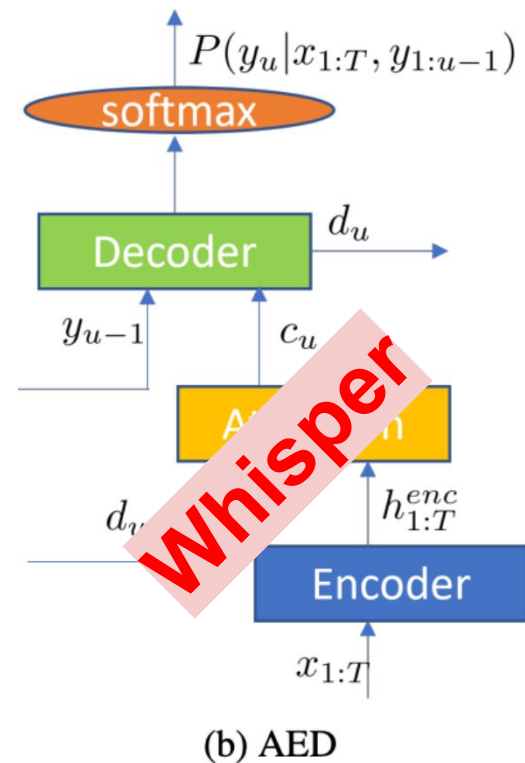
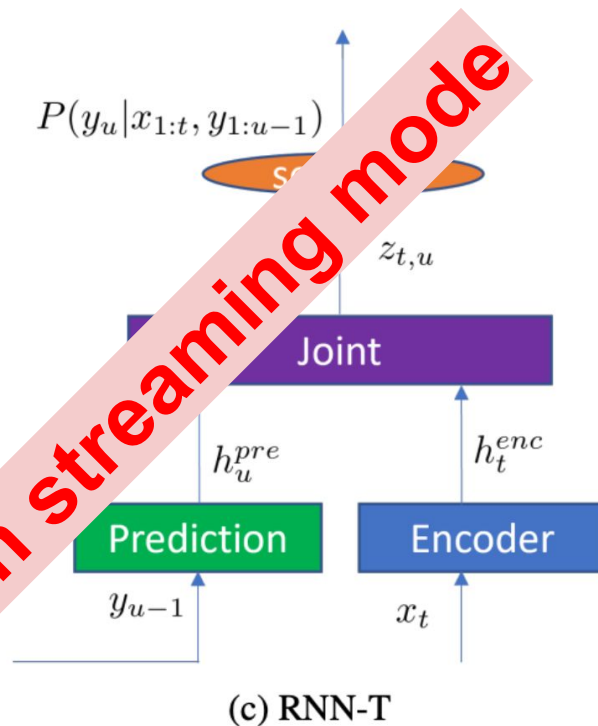
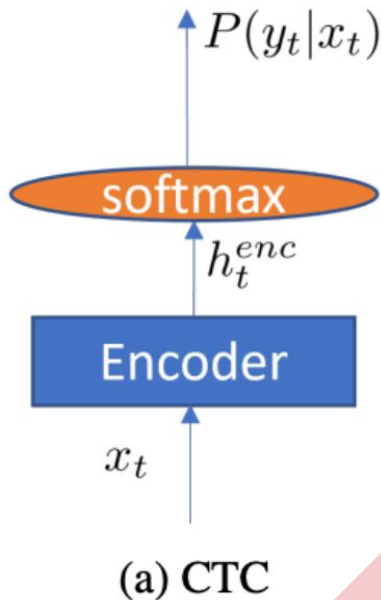


Fig. 1. The HuBERT approach predicts hidden cluster assignments of the masked frames (y₂, y₃, y₄ in the figure) generated by one or more iterations of k-means clustering.





Can work in streaming mode

Whisper

Name (main driver)	Community	Neural network backend	Streaming ASR supported	Hybrid / E2E	License	Reference
Kaldi (JHU, D. Povey)	Large but decreasing	Kaldi	YES	Hybrid	Apache 2.0	https://kaldi-asr.org/
ESPnet (CMU, S. Watanabe)	Large	PyTorch	YES (partially)	Hybrid, E2E	Apache 2.0	https://github.com/espnet/espnet
k2-fsa (Xiaomi, D. Povey)	Small but increasing	PyTorch	YES	E2E	Apache 2.0	https://github.com/k2-fsa/k2 (training) https://github.com/k2-fsa/sherpa (runtime)
SpeechBrain (MILA, M. Ravanelli)	Medium	PyTorch	NO	E2E	Apache 2.0	https://github.com/speechbrain/speechbrain
NeMo (Nvidia)	Large	PyTorch Lightning	YES (partially)	E2E	Apache 2.0	https://github.com/NVIDIA/NeMo
Athena (Beike AAC, Beijing)	Small and decreasing	TensorFlow	YES	E2E	Apache 2.0	https://github.com/athena-team/athena
Espresso (JHU, Yiming Wang)	Large but decreasing	PyTorch	NO	E2E	MIT License	https://github.com/freewym/espresso
WeNet (Horizon Robotics, B. Zhang)	Medium and increasing	PyTorch	YES	E2E	Apache 2.0	https://github.com/wenet-e2e/wenet



Parakeet model

- Příprava dat (lhotse manifesty)
- Výběr modelu
- Nastavení hyper parametrů
 - WandB Sweeps
- Nastavení logování
 - WandB (TensorBoard)
- Implementace scorovací metriky
- (Implementace loss funkce)

A pak už jen získat nějaké GPU a pustit to.

Škálování (trénovaná na HPC) ...

https://github.com/xnevar00/ASR-demo/blob/main/parakeet_finetune_demo.ipynb



https://bit.ly/czai_asrdemo

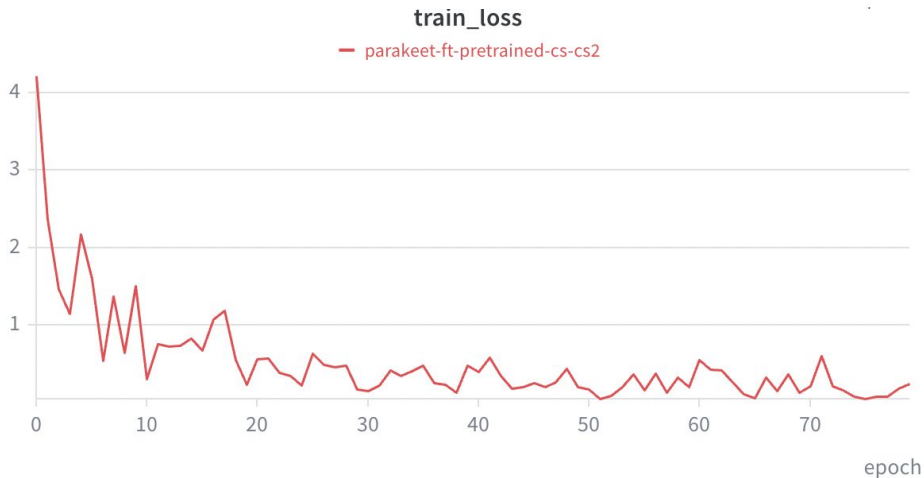
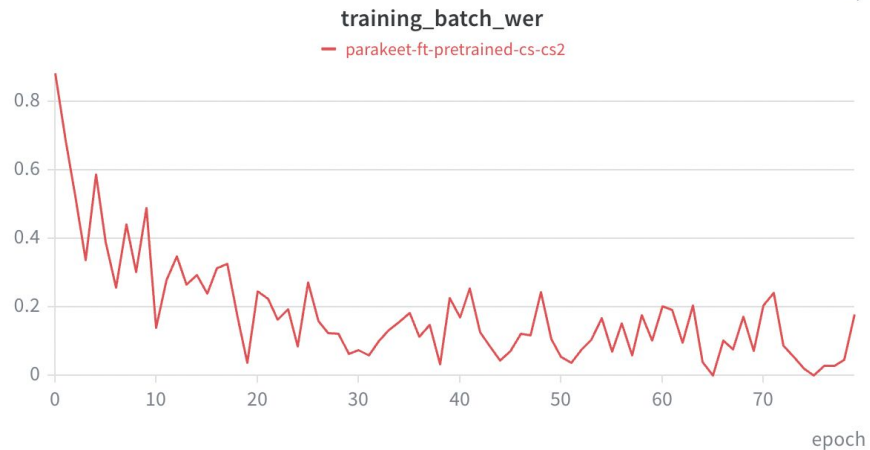
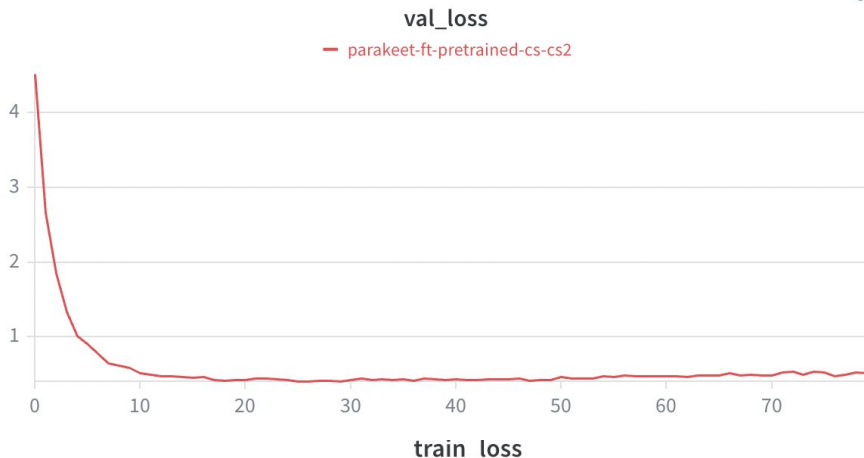
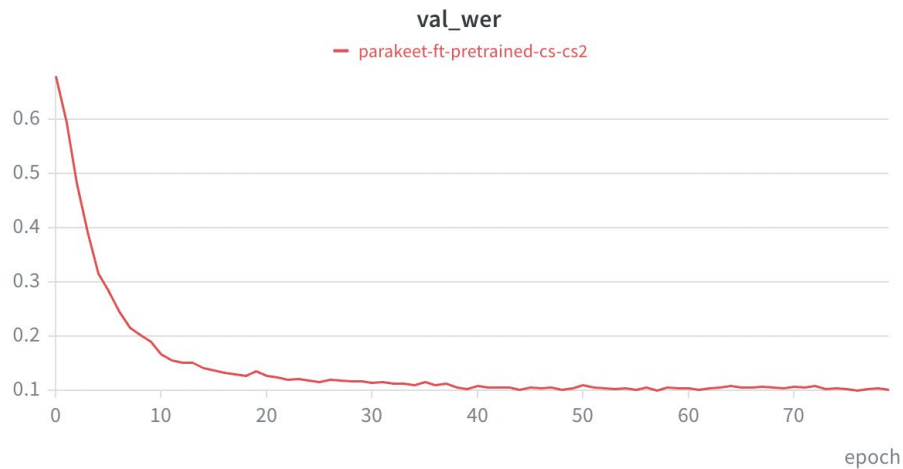
```
train_ds_cfg = OmegaConf.create({
    "cuts_path": TRAIN_CUTS,
    "sample_rate": 16000,
    "batch_size": 1,
    "shuffle": True,
    "num_workers": 1,
    "use_lhotse": True,
    "min_tps": -1,
    "max_tps": float("inf"),
    "pretokenize": False,
})

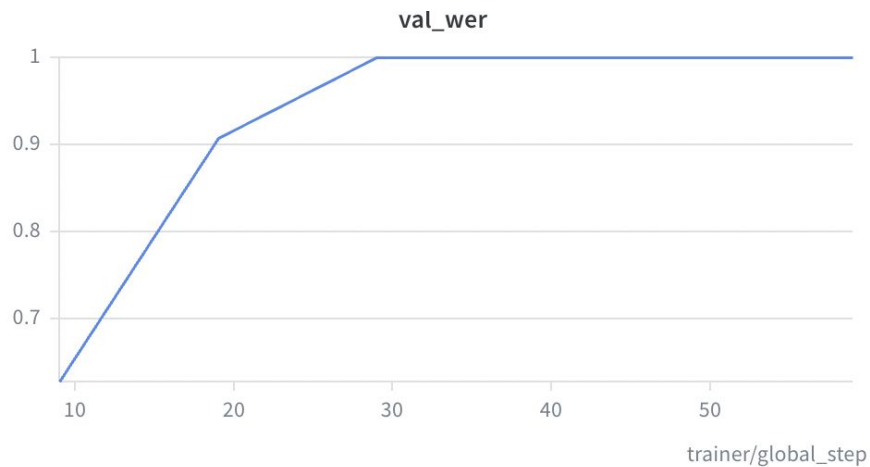
validation_ds_cfg = OmegaConf.create({
    "cuts_path": VAL_CUTS,
    "sample_rate": 16000,
    "batch_size": 1,
    "shuffle": False,
    "num_workers": 1,
    "use_lhotse": True,
    "min_tps": -1,
    "max_tps": float("inf"),
    "pretokenize": False,
})
```

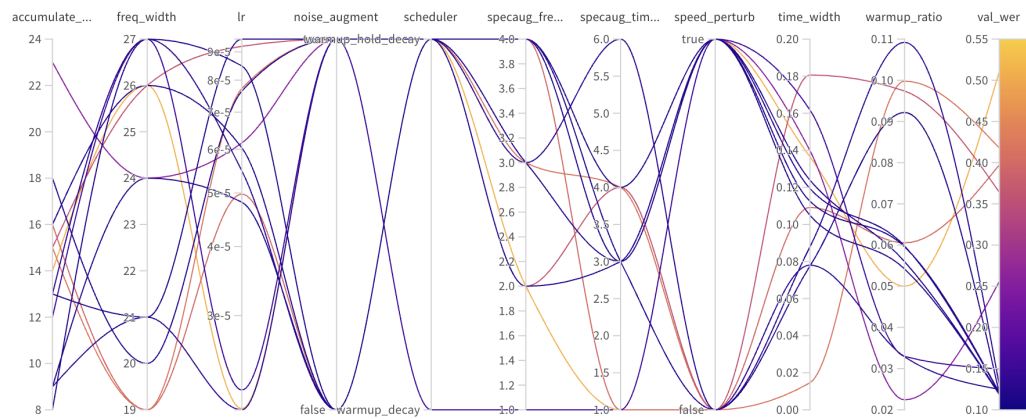
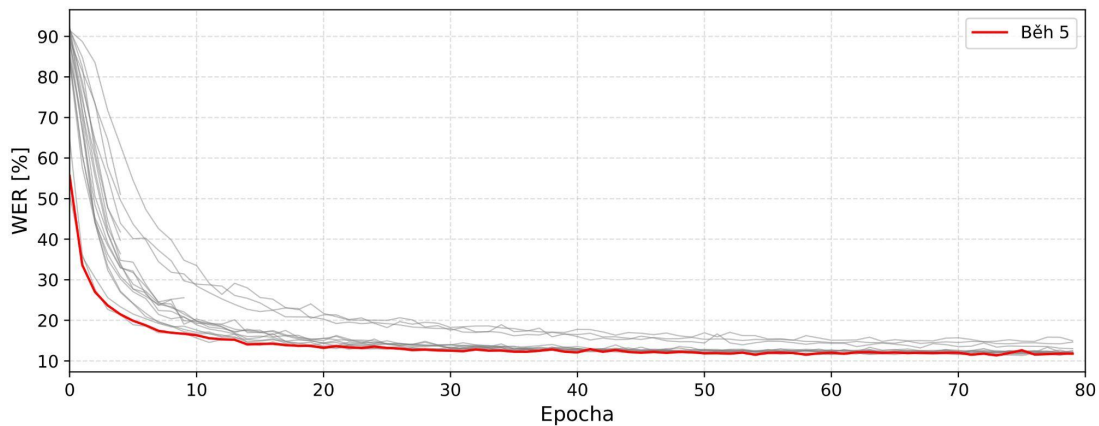
```
with open_dict(asr_model.cfg):
    asr_model.cfg.optim = OmegaConf.create({
        "name": "adamw",
        "lr": 1e-4,
        "betas": [0.9, 0.98],
        "weight_decay": 0.0,
    })
asr_model.setup_optimization(asr_model.cfg.optim)

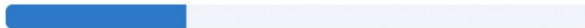
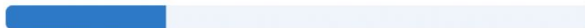
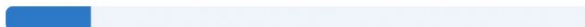
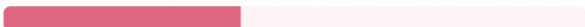
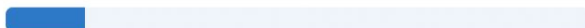
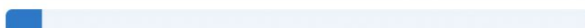
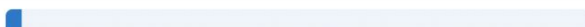
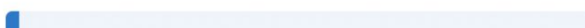
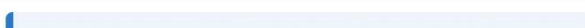
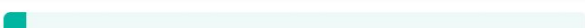
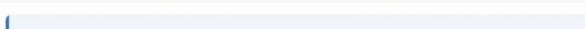
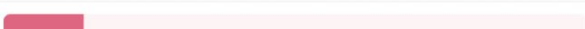
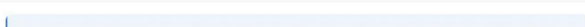
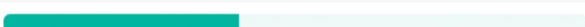
trainer = pl.Trainer(
    accelerator="gpu" if torch.cuda.is_available() else "cpu",
    devices=1,
    strategy="auto",
    max_epochs=7,
    precision="bf16-mixed" if torch.cuda.is_available() else "32-true",
    val_check_interval=1.0,
    log_every_n_steps=1,
    enable_checkpointing=False,
    logger=wandb_logger,
)
```

```
asr_model.set_trainer(trainer)
trainer.fit(asr_model)
```







Config parameter	Importance ⓘ ↓	Correlation
scheduler.value_warmup_hold_decay		
scheduler.value_warmup_decay		
specaug_time_masks		
time_width		
specaug_freq_masks		
lr		
warmup_ratio		
speed_perturb		
Runtime		
noise_augment		

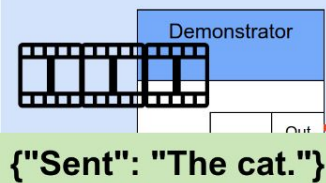
Runtime (inference)

Deployment a (dlouhodobý) runtime může být úplně jiná story.

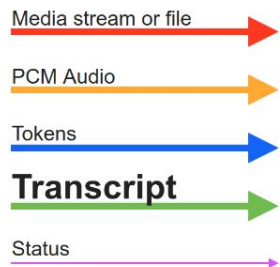
- robustnost (memory leaks, dlouhé nahrávky, poškozené nahrávky, ...)
- škálování
- streaming

Preprocessing / Postprocessing?

Pro náš účel je Python script OK :)

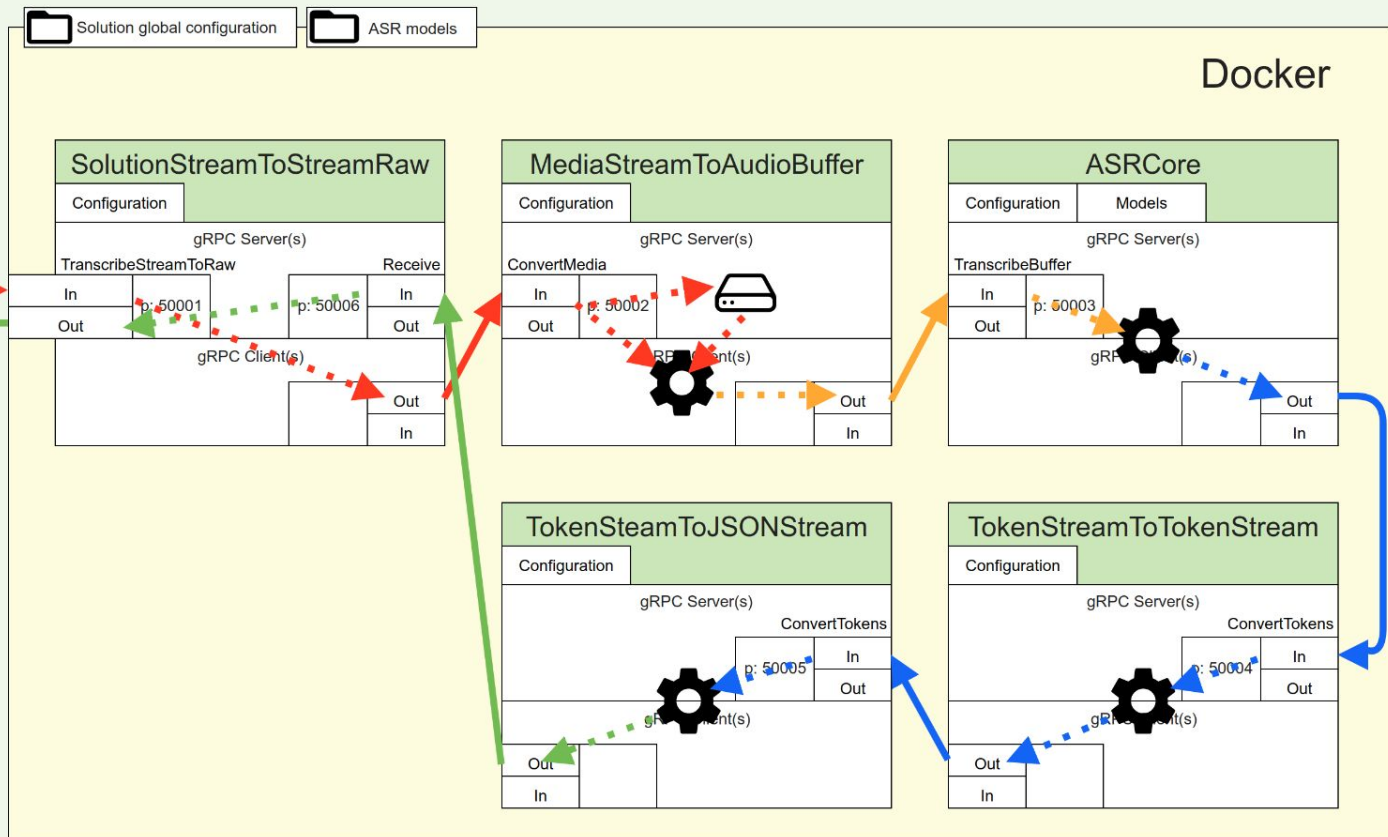


Legend



Visualisation of module internal data flow

p: 50001 means TCP/IP port # 50001



<https://pcspeech-demo.fit.vut.cz/atc/>

BUT Speech@FIT ATC VHF demonstrator

Demonstrator of automated speech transcription of ATC communication developed by [BUT Speech@FIT](#) research group. This work is based on projects [ATCO2](#), [HAWAII](#) and other. The underlying model is finetuned NeMo Parakeet model. Big thanks to Veronika Nevarilova for finetuning the model. Contact [Igor Szoke](#) for details or read about other [projects](#) or go back to other [demos](#).

Stop listening Prague Airport (LKPR) Tower at 134.560 MHz

Audio connected

Live Waveform

10s window | peak -4.2 dBFS | 8000 Hz



Latest Utterances

9/2/2026, 2:55:11 PM | 2.51s | SNR 21.0 dB

 tower good afternoon ryanair two whiskey papa established

9/2/2026, 2:54:16 PM | 2.56s | SNR 26.0 dB

 contacting praha radar approaching five zero ahoj ?? hi/bye ??

9/2/2026, 2:54:13 PM | 2.14s | SNR 56.0 dB

 private wings eight zero contact Praha Radar

9/2/2026, 2:53:14 PM | 1.88s | SNR 41.0 dB

 Praha Radar ryanair nine foxtrot foxtrot

9/2/2026, 2:53:09 PM | 2.62s | SNR 48.0 dB

 Ryanair nine foxtrot foxtrot contact Praha Radar hezký den

Děkujeme za
pozornost

szoke@fit.vut.cz

inevarilova@fit.vut.cz