

# Live Credible Translation

**Dominik Macháček**

23.5.2025

[machacek@ufal.mff.cuni.cz](mailto:machacek@ufal.mff.cuni.cz)

## Live Credible Translation

I will introduce my project “Live Credible Translation.” The current automatic Simultaneous Speech Translation (SST) systems usually provide only one translation hypothesis without any Quality Estimation (QE) score that would indicate how likely is the translation correct or wrong. I plan to propose SST QE methods for practical applications, such as for SST of cross-lingual dialogues, and for human real-time post-editing of SST.

First, I will introduce my current work, (1) a state-of-the-art simultaneous translation system, and (2) a corpus of cross-lingual dialogues. Then, I present my plans and ask for recommendations.

# Outline

- Who am I
- Live Credible Translation
  - Intro
  - Brief Plan
- SoTA Simultaneous Speech Translation
- Cross-Lingual Dialogues
- Plans+Discussion

# My main interests

- **Simultaneous Speech Translation**, Simultaneous Interpreting, MT, ASR
- Technology transfer to industry:

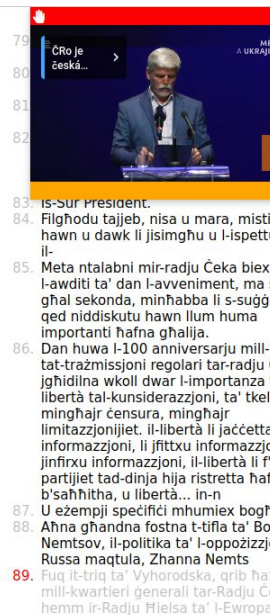
## ELITR “AI Interpreting” as a service

Contact me for live demo!

- Postdoc project:

## Live Credible Translation

- powerful experience.
80. That was Josef Pazderka from Czech Radio Plus.
81. And now, the President of the Czech Republic, Petr Pavel.
82. Please come up here on stage, and present your opening speech to start the first session of this conference.
83. conference, Ukraine as a Shared Responsibility.
84. Mr. President.
85. Good morning, ladies and gentlemen, guests here and listeners and viewers on the other platforms.
86. When I was asked by the Czech radio to take over the auspices of this event, I did not hesitate for a second, because the topics that we are discussing here today are very important to me.
87. This is the 100th anniversary since the start of the regular broadcast of the Czech radio, which also tells us about the importance of freedom of speech, of talking without censorship, without limitations, the freedom to accept information, to seek information, to spread information, the freedom that in many parts of the world is restricted very strongly, and a freedom... people keep giving their lives for.
88. And specific examples are not far away.
89. We have among us the daughter of Boris Nemtsov, the murdered Russian opposition politician, Zhanna Nemtsova.
90. On Vyhorodska street, quite close to the headquarters of the Czech Radio, there is Radio Free Europe, and three of its journalists are now in prison,
91. republiky, Petra Pavla, aby přišel sem k nám a přednesl svůj úvodní projev a vlastně tak otevřel ten první blok celé konference.
92. Blok nazvaný Ukrajina jako společná odpovědnost.
93. Prosim, pane prezidente.
94. Dobry den, damy a panove, vazeni hoste zde v sale, posluchaci, ale take divaci na ostatnich platformach.
95. Kdyz me vedeni Ceskeho rozhlasu pozadalo o zastitu nad dnešní konferenci, nemusel jsem dlouho váhat, protože témata, kterými se tady zabýváme, jsou pro mě velice důležitá.
96. Připomínáme si 100. výročí odzahžení pravidelného rozhlasového vysílání a to je zároveň i připomínkou významu svobody slova.
97. Svobody vyjadřovat se bez cenzury a bez omezení.
98. Svobody přijímat informace a myšlenky, vyhledávat je a šířit.
99. Svobody, která je v různých koutech světa stále výrazně omezována a za její šprosazování lidé i dnes platí tu nejvyšší cenu.
100. Pro konkrétní příklady nemusíme vůbec chodit daleko.
101. Mezi námi je dnes dcera zavražděného ruského opozičního politika Borise Němcova, žena Němcovová.
102. Na ulici Vinohradská, jen kousek od sídla Českého rozhlasu, sídlí i Radio Sobotná Evropa.
103. Jehož tři novináři jsou dnes vězněni. – Jiřard Losik a Andrej Kuzněčik v Bělorusku a Vladislav Jesipenko na ruském okupovaném Krymu.
104. V únoru tohoto roku jsme si připomněli pět let od vraždy slovenského

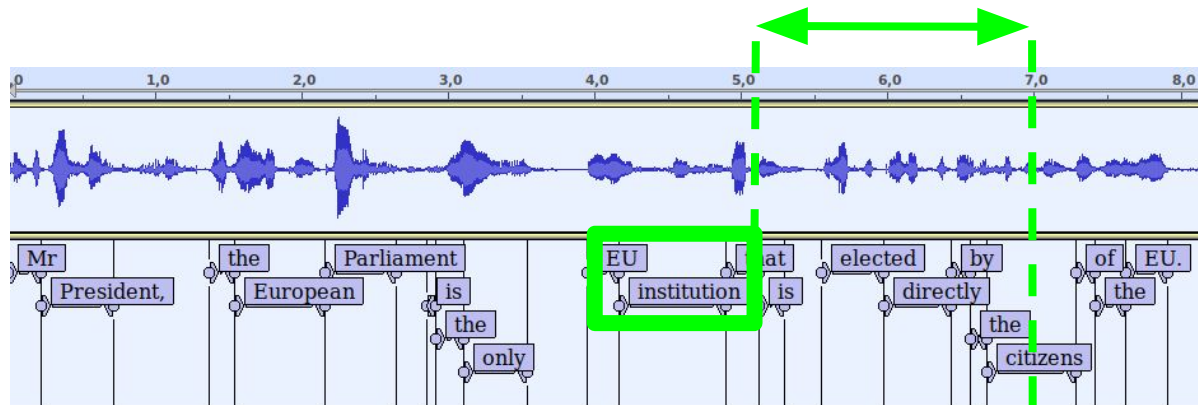


**Live Credible Translation**

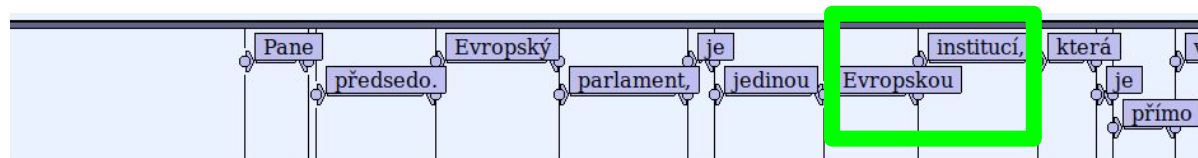
# "Live" = Simultaneous Speech Translation

short additive **delay**

... 2-4 seconds (avg.)



Original source



**Simultaneous**  
Speech Translation

# Sim. Speech Translation $\neq$ Automatic Interpreting

Simultaneous Speech Translation  $\subseteq$  Simultaneous Interpreting

Simultaneous

Translating speech

Simultaneous

Translating speech when needed

Summarization

Simplification

Intercultural transfer

Paralinguistics

“Human intelligence:” ethics, responsibility, ...

# Sim. ST challenges we focus on:

## ❑ Technical

- ❑ Speech acquisition
- ❑ ASR, speech processing
- ❑ HW, fast computation

## ✓ **Linguistics**

- ❑ How to translate?
- ❑ Simultaneity: Wait or translate?

## ❑ Practical deployment

- ❑ Cost
- ❑ User interface
- ❑ Instructing the users
- ❑ Managing expectation



# "Credible" Simultaneous Speech Translation

	<b>SoTA</b> *Problems in current translations:	<b>Beyond SoTA</b> We aim to improve in 2025-2027:	<b>Use in application</b> Example future work:
<b>Acoustics</b> Noisy audio, ...	*Titulky vytvořil JohnnyX. (noise often creates "hallucination," nonsense output)	<b>Noise/hallucination detection:</b> Titulky vytvořil JohnnyX.	<b>Reject:</b> Sorry, the translation is not available. The audio is too noisy.
<b>Speech processing</b> Non-standard accent, ...	*Lascaux, Dava, Cridla. (French L1 speaks L2 Czech. En translation expected: Love gives wings.)	<b>Estimate speech recognition quality in translation:</b> Lascaux, Dava, Cridla.	<b>Delayed but more correct:</b> Sorry, the accent is challenging, processing longer..... Love gives wings.
<b>Translation quality</b>	If you have bubble *fuchs, you can pull to pull the *spalier (Czech src: Pokud Máte bublifuk, můžete ho použít ve špalíru.)	<b>Quality Estimation:</b> If you have bubble fuchs, you can pull to pull the spalier.	<b>Live post-editing:</b> If you have a bubble blower, you can pull it in the espalier.
<b>Simultaneity</b>	Doctor met the patient = Doktor* se setkal* s pacientem [pause] that she cured. = , kterého vyléčila. (Future content can change translation – the doctor's gender.)	<b>Uncertainty Estimation:</b> Doktor/ka se setkal/a s pacientem  , kterého vyléčila.	<b>Explain in natural language:</b> Doktor (ale mohla by to být taky doktorka) se setkal s pacientem, kterého vyléčil. (Aha, tak je to doktorka.)

# “Practical” Quality Estimator for SST

	<b>SoTA Limitations</b>	<b>Beyond SoTA</b> We aim to improve in 2025-2027:
<b>Efficiency</b>	large Deep Learning model for SST + second large one for QE would be <b>too costly</b>	<b>Efficient QE:</b> Unsupervised methods, small models, adapters
<b>Explainability</b>	QE score too vague, not explainable	Confidence intervals, detect origins of errors (acoustics vs. speech processing vs. translation vs. simultaneity)
<b>Reliability</b>	No or not reliable benchmarks	Benchmark vs. real-life evaluation, human eval.

## 1. Acquire baselines

- Sim. ASR + Sim. ST
- QE for ASR
- QE for MT

## 2. Benchmark


- Evaluation data
  - Language pairs
  - Domain, use case, ...
- Evaluation method

## 3. Improvements

# Plan + Status

## 1. Acquire baselines

- Sim. ASR + Sim. ST
- QE for ASR
- QE for MT

...   
... seems close  
... in next 2 months?

## 2. Benchmark

- Evaluation data
  - Language pairs
  - Domain, use case, ...
- Evaluation method

... ongoing, but good enough?

... should be simple

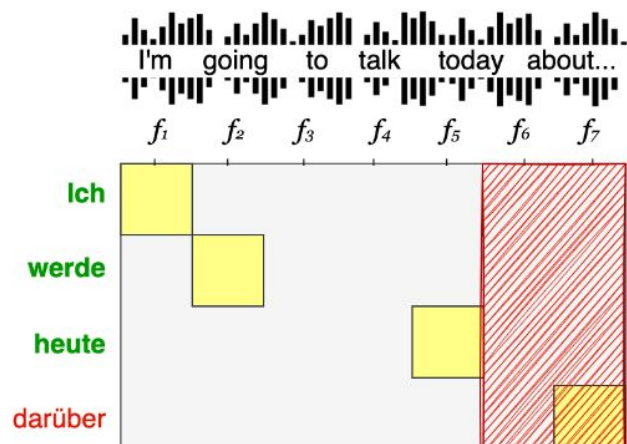
## 3. Improvements

... **ideas**

**State-of-the-art SST system**

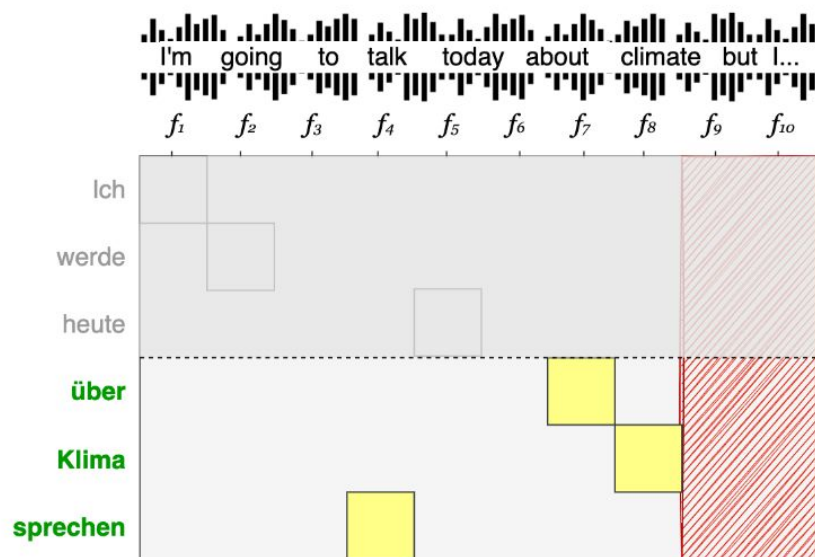
# State-of-the-art SST system

- Submission to IWSLT 2025 Simultaneous Shared Task
- Czech->English, English->{German, Japanese, Chinese}
- Whisper model + AlignAtt simultaneous policy



# State-of-the-art SST system

- Submission to IWSLT 2025 Simultaneous Shared Task
- Czech->English, English->{German, Japanese, Chinese}
- Whisper model + AlignAtt simultaneous policy



# State-of-the-art SST system

- Submission to IWSLT 2025 Simultaneous Shared Task
- Czech->English, English->{German, Japanese, Chinese}
- Whisper model + AlignAtt simultaneous policy

- + beam search => n-best list => QE for ASR

1. to the Chamber of Deputies
2. to the Assembly

- + prompt for terminology

poslanecká sněmovna =  Chamber of Deputies  Senate

- + context between processing units

<|startofprev|>We are preparing it and it will of course be submitted to<|startoftranscript|><|cs|><|translate|><|notimestamps|> the Chamber of Deputies



# State-of-the-art SST system

- Whisper ASR to English + EuroLLM into De, Ja, Zh + LocalAgreement

- System prompt:

You are conference interpreter into {TGT-Language}.

- In-context example:

Please, go ahead, you can start with your presentation, we are ready. Hello,  
Bitte schön, Sie können mit Ihrer Präsentation beginnen, wir sind bereit. Guten Tag,

- Inter-sentence context:

300 tokens = appx. 43 seconds (but maximum is 11 minutes!)

SRC Hello, this is Jiawei Zhou from Harvard University. I am very glad to present our work on Online Semantic Parsing for Latency Reduction in Task-Oriented Dialogue. This is joint work with Jason, Michael, Anthony and Sam from Microsoft Semantic  
TGT Guten Tag, hier ist Jiawei Zhou von der Harvard University. Ich freue mich sehr, unsere Arbeit zum Thema Online Semantic Parsing zur Reduzierung der Latenz in aufgabenorientierten Dialogen vorzustellen. Dies ist eine gemeinsame Arbeit mit Jason, Michael, Anthony

# Results

- Czech-to-English
  - Works both for 2s and 4s latency
  - Dev: improvement over 2 BLEU over baseline
  - Test: winning 3 cases out of 4 (preliminary)
    - Human eval pending
- English-to-{German,Chinese,Japanese}
  - Works only with the 4-5 seconds latency
  - Dev: 8-22 BLEU improvements over baseline
  - Test: De+Ja winner, Zh second (preliminary)
    - Human eval pending

# Limitations, summary

- Strengths:
  - Works very well
  - Strong baseline for next work
  - Very multilingual: 99 src -> English -> 35 tgt
- Limitations:
  - Hallucinations!
  - Not real-time simulation
  - Not many submissions
  - Unseen and noisy acoustics, dialogues
  - Non-standard language: accents, casual and spontaneous speech
  - Prompt and context may help or hurt

## **Corpus of cross-lingual dialogues**

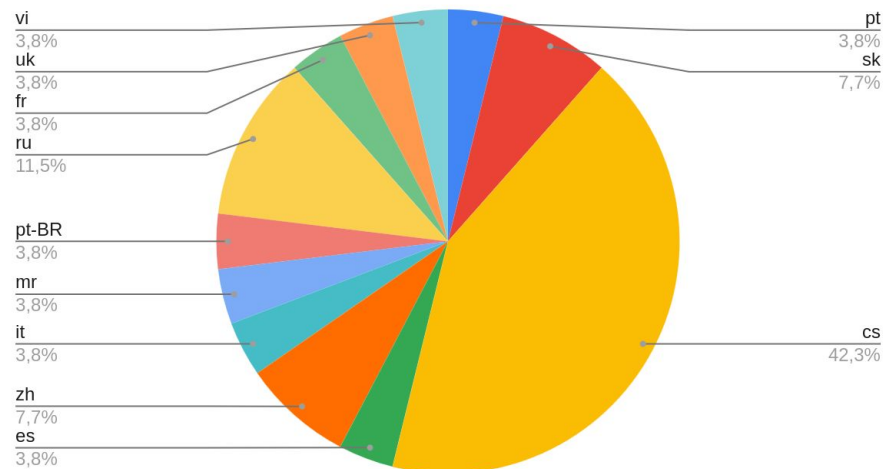
# InCroMin: Corpus of cross-lingual dialogues

- Dialogues of 2-5 ppl without common language, mediated by SST
- Some “genuine information need” (to ask or to tell)
- Topics freely chosen by participants.
  - Acted interview of a refugee and an officer in an integration center (2)
  - Professor-student consultation (several)
  - Work meetings (several)
  - Chitchats
- Content (mostly)
  - Audio
  - ASR + MT into English
  - Corrected ASR + MT => **useful for QE**
  - Pseudonymization
  - ...
- Consents to publish (mostly)

# InCroMin: Status

- Publishable: ~21 meetings á 30min, 16 languages, many speakers
- Languages (mostly):
  - No non-English speakers
- Processing in progress
- **New volunteers welcome!**

Count of Languages



## **Improvements + Discussion**

# Plans/Ideas for Improvements + Discussion

- Whisper ASR confidence via n-best list [K. Beneš dissertation, BUT 2025]
  - Then apply to speech transl. via attention
- Unsupervised MT and ST QE
- Severity of errors in Simultaneous Speech Translation:
  - Fluency errors are negligible for understanding ... přípytek
  - Misspelled name is OK when it's close and the partner knows ... Lucí Borhiovou
  - What is a “**catastrophic error**” in simultaneous interpreting?
  - => target them primarily
- Tips on **primary use cases** of Live Credible Translation?
- **Multilinguality**: Experiment with more than 3 language pairs? Why yes/no?
- Linguistic approach:
  - Model phases of simultaneous interpreting.
  - Specify one subproblem, create a QE for that using a linguistic knowledge or a resource.
  - Combine many such QEs => SST QE can detect origin of an error.
  - What **linguistic knowledge** or a **resource** could be useful?
    - Cognates, loan words, false friends, L2 corpora (spoken, written)?
    - Shared knowledge of the communication background: how to define/categorize it?



# Phases of simultaneous interpreting

- Model phases of simultaneous interpreting / speech-to-text translation
  - Speaker: intention
  - Sound acquisition
  - Src. speech detection
  - Src. language detection
  - Src. acoustic model
  - Src. language model
  - Translation: src analysis -> transfer -> tgt synthesis
  - Receiver: intention
  - Receiver: perception
    - Src. + tgt. vs. tgt only?
    - Modalities: src. speech + vision, tgt text vs. speech
    - Background info: communication setup, personal experience, education, ...
    - Src. language knowledge (+ tgt. language knowledge)
- Apply methods for specific subproblems, combine them