On Speech Recognition Model Adaptation and Accuracy Calibration for Oral History Archives

Practical routes from one recording to an archive-scale workflow

Jan Švec Jan Lehečka Pavel Ircing

Department of Cybernetics, University of West Bohemia in Pilsen

DHECC 2026, Odense 24 September 2026

Search begins with a transcript

Search and thematic querying

Scholars can locate relevant passages across collections that are too large for serial listening.

Verification remains essential

Names, places, and quotations still require listening against the original recording.

A demanding oral-history collection

MALACH

Holocaust survivor testimonies

USC Shoah Foundation Visual History Archive

Why generic models struggle

Older speakers, emotional and spontaneous speech, non-native accents, and uneven recordings challenge generic recognisers.

2,307 h

Training material

used for model adaptation

English	246 h
German	1,803 h
Czech	87 h
Slovak	96 h
Polish	52 h
Hungarian	23 h

Source Paper Table 1; Visual History Archive, vha.usc.edu

Specialisation helps unevenly across languages

Word accuracy	EN	DE	CS	SK	PL	HU
Generic model	88.54	86.85	85.98	89.72	84.32	**83.65**
Oral-history model	**89.83**	**88.44**	**92.68**	**89.72**	**84.43**	83.36
Change	+1.29	+1.59	+6.70	0.00	+0.11	-0.29

The largest gains occur where generic training already covered diverse speech.

Source Word accuracy on the oral-history test sets. Values in percentage points; paper Table 2.

A recording-level estimate supports practical decisions

80%

estimated word accuracy

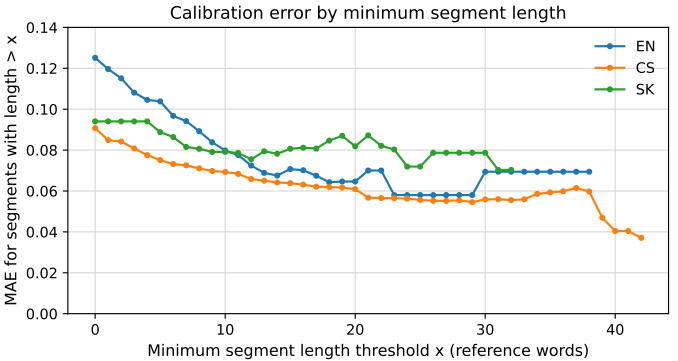
Useful for triage

Index a strong transcript, plan manual correction, or direct attention to a difficult recording.

An estimate, not a guarantee

Important passages and quotations still need verification against the audio.

Reliability estimates improve over longer passages



Whole recordings

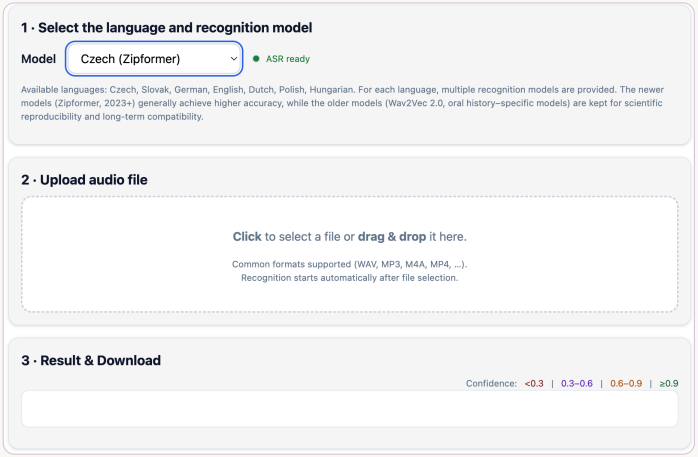
Across all languages, average error stayed below five percentage points.

Shorter passages

Use at least 30 words. Short fragments give unstable estimates.

Source Speaker-disjoint evaluation with clean and noise-augmented test recordings; paper Table 3 and Figure 4.

Transcription and reliability in one service



Six oral-history models

English, German, Czech, Slovak, Polish, and Hungarian.

Several outputs

Plain text, subtitles, timed JSON, and Transcriber XML.

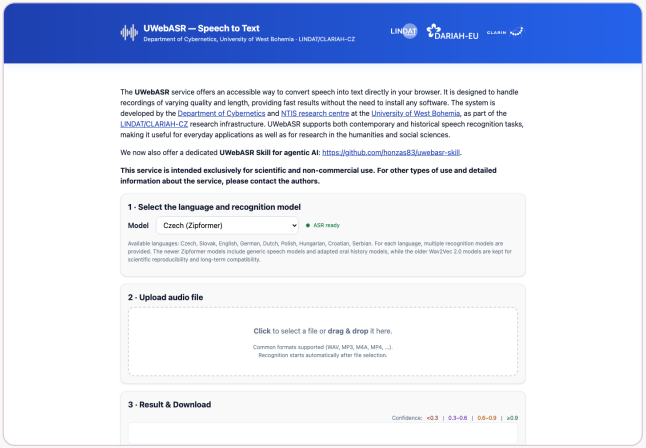
One visible estimate

Expected accuracy when recognition finishes.

Five ways to use the same recogniser

The right route depends on the user, collection size, and governance requirements.

1. Web interface for a single recording



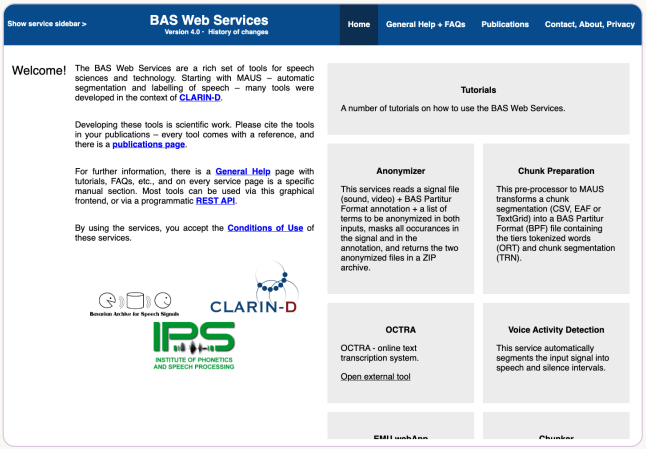
No installation

Upload one recording, inspect the result, and export it.

Open UWebASR

uwebasr.zcu.cz

2. Bavarian Archive for Speech Signals (BAS)



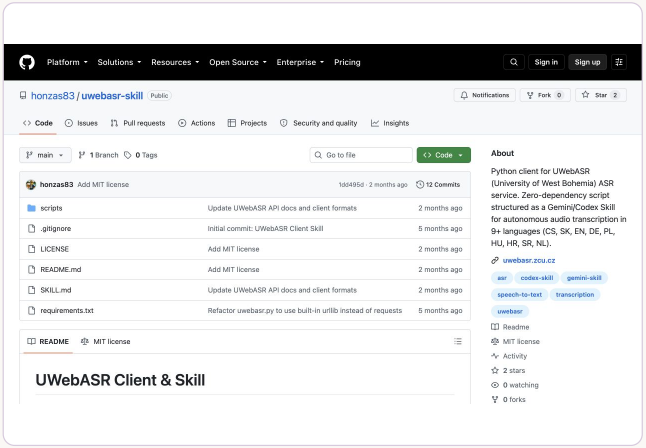
Within CLARIN

OCTRA supports correction, alignment, and annotation export.

BAS Web Services

clarin.phonetik.uni-muenchen.de

3. An AI assistant handles routine transcription



Ordinary-language request

"Transcribe every recording in this folder."

Get the skill

`github.com/honzas83/uwebasr-skill`

4. APIs connect UWebASR to existing systems

OpenAI-compatible endpoint

The same request shape as OpenAI Speech-to-Text.

POST `/v1/audio/transcriptions`

`uwebasr.zcu.cz`

```
model   malach/en/zipformer
file    interview.wav
```

No API key required.

Direct REST access

Return transcripts and confidence scores.

API documentation

`uwebasr.zcu.cz/#h-api`

5. Offline HPC supports entire collections

33,902
testimonies processed

60,000
hours of recorded speech

27 TB
source video

Containerised deployment

The same recogniser stack can run in controlled infrastructure through scheduled parallel jobs.

Reproducible processing

A fixed container and model version can support reruns and long-term projects.

Source Brückner, Lehečka, Švec and Pecina (2026), *Modeling the Language of Holocaust Survivors' Testimony with Domain-Adapted Transformers*, HTRes @ LREC 2026.

Privacy requirements influence the access route

Public UWebASR service

Audio and generated transcripts are processed in memory and discarded when the session finishes.

Processing runs on physical University of West Bohemia servers in Plzeň.

Controlled deployment

Restricted collections can use an institutionally approved offline HPC environment.

Calibration can keep reference transcripts local; only audio reaches the selected online recogniser.

Collection policy remains the deciding authority.

Source UWebASR data handling and privacy policy: uwebasr.zcu.cz/#h-privacy

Choosing an access route

ROUTE	BEST FIT	DATA HANDLING
Web	One recording	Transient UWebASR processing
BAS Web Services	Annotation project	Selected CLARIN service policy
Agent	Files and folders	Host and endpoint policies apply
API	System integration	Transient UWebASR processing
Offline HPC	Large collections	Institution-controlled processing

Restricted collections usually require institution-controlled processing.

Searchable collections with visible reliability

Specialised oral-history models are available now

Six languages, several output formats, and a recording-level estimate of transcript accuracy.

Access can match the institution

Use a browser, an established portal, an API, an assistant, or offline infrastructure.

Questions?

Try UWebASR

uwebasr.zcu.cz