

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.

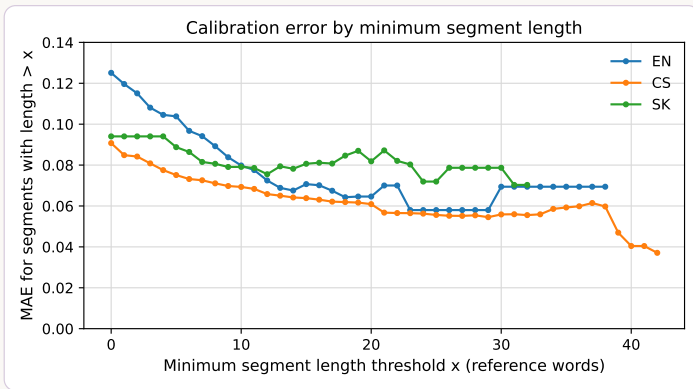
Reliability estimates improve over longer passages

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

1 · Select the language and recognition model

Model Czech (Zipformer) ▾ ● ASR ready

Available languages: Czech, Slovak, German, English, Dutch, Polish, Hungarian. For each language, multiple recognition models are provided. The newer models (Zipformer, 2023+) generally achieve higher accuracy, while the older models (Wav2Vec 2.0, oral history-specific models) are kept for scientific reproducibility and long-term compatibility.

2 · Upload audio file

Click to select a file or **drag & drop** it here.

Common formats supported (WAV, MP3, M4A, MP4, ...).
Recognition starts automatically after file selection.

3 · Result & Download

Confidence: <0.3 | 0.3–0.6 | 0.6–0.9 | ≥0.9

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

 **UWebASR — Speech to Text**
Department of Cybernetics, University of West Bohemia - LINDAT/CLARIAH-CZ

The **UWebASR** service offers an accessible way to convert speech into text directly in your browser. It is designed to handle recordings of varying quality and length, providing fast results without the need to install any software. The system is developed by the [Department of Cybernetics](#) and [NTIS research centre](#) at the [University of West Bohemia](#), as part of the [LINDAT/CLARIAH-CZ](#) research infrastructure. UWebASR supports both contemporary and historical speech recognition tasks, making it useful for everyday applications as well as for research in the humanities and social sciences.

We now also offer a dedicated **UWebASR Skill for agentic AI**: <https://github.com/honzas83/uwebasr-skill>.

This service is intended exclusively for scientific and non-commercial use. For other types of use and detailed information about the service, please contact the authors.

1 · Select the language and recognition model

Model Czech (Zipformer) ASR ready

Available languages: Czech, Slovak, English, German, Dutch, Polish, Hungarian, Croatian, Serbian. For each language, multiple recognition models are provided. The newer Zipformer models include generic speech models and adapted oral history models, while the older Wav2Vec 2.0 models are kept for scientific reproducibility and long-term compatibility.

2 · Upload audio file

Click to select a file or drag & drop it here.

Common formats supported (WAV, MP3, M4A, MP4, ...).
Recognition starts automatically after file selection.

3 · Result & Download

Confidence: <0.3 | 0.3-0.6 | 0.6-0.9 | ≥0.9

No installation

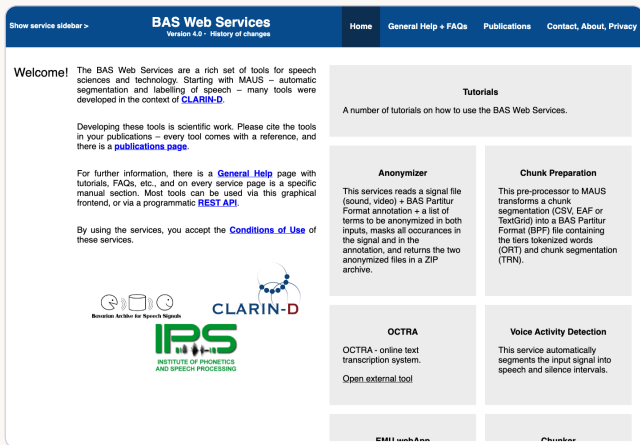
Upload one recording, inspect the result, and export it.



Open UWebASR

uwebasr.zcu.cz

2. Bavarian Archive for Speech Signals (BAS)



The screenshot shows the homepage of the BAS Web Services. The header is dark blue with the text "BAS Web Services" and "Version 4.0 - History of changes". Below the header is a navigation bar with links: "Home", "General Help + FAQs", "Publications", and "Contact, About, Privacy". The main content area is white and contains a "Welcome!" message, a "Tutorials" section, and a grid of service tiles. The "Tutorials" section mentions that a number of tutorials are available on how to use the BAS Web Services. The grid includes tiles for "Anonymizer", "Chunk Preparation", "OCTRA", and "Voice Activity Detection". At the bottom left, there are logos for "Bavarian Archive for Speech Signals", "CLARIN-D", and "IPS INSTITUTE OF PHONETICS AND SPEECH PROCESSING".

Show service sidebar > **BAS Web Services**
Version 4.0 - History of changes

Home General Help + FAQs Publications Contact, About, Privacy

Welcome! The BAS Web Services are a rich set of tools for speech sciences and technology. Starting with MAUS – automatic segmentation and labelling of speech – many tools were developed in the context of [CLARIN-D](#).

Developing these tools is scientific work. Please cite the tools in your publications – every tool comes with a reference, and there is a [publications page](#).

For further information, there is a [General Help](#) page with tutorials, FAQs, etc., and on every service page is a specific manual section. Most tools can be used via this graphical frontend, or via a programmatic [REST API](#).

By using the services, you accept the [Conditions of Use](#) of these services.

Tutorials
A number of tutorials on how to use the BAS Web Services.

Anonymizer
This services reads a signal file (sound, video) + BAS Partitur Format annotation + a list of terms to be anonymized in both inputs, masks all occurrences in the signal and in the annotation, and returns the two anonymized files in a ZIP archive.

Chunk Preparation
This pre-processor to MAUS transforms a chunk segmentation (CSV, EAF or TextGrid) into a BAS Partitur Format (BPF) file containing the tiers tokenized words (ORT) and chunk segmentation (TRN).

OCTRA
OCTRA - online text transcription system.
[Open external tool](#)

Voice Activity Detection
This service automatically segments the input signal into speech and silence intervals.

CLARIN-D
IPS
INSTITUTE OF PHONETICS AND SPEECH PROCESSING

Within CLARIN

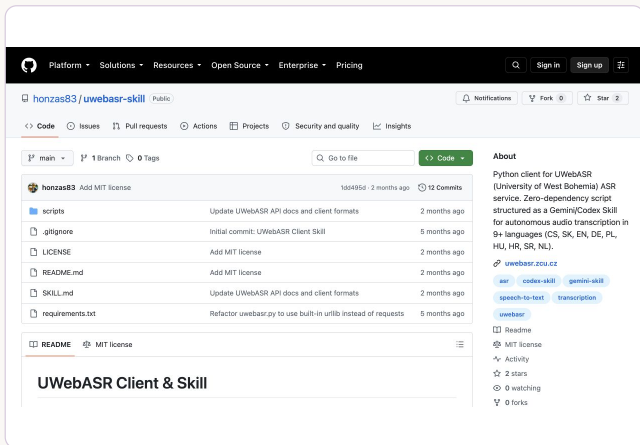
OCTRA supports
correction, alignment,
and annotation export.



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](https://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