Journal of Behavioral and Experimental Finance 41 (2024) 100891

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Finance

journal homepage: www.elsevier.com/locate/jbef

Ready-to-use software article

Take the aTrain. Introducing an interface for the **Ac**cessible **Tra**nscription of **In**terviews

Armin Haberl [a,*], Jürgen Fleiß [a], Dominik Kowald [b], Stefan Thalmann [a]

[a] *Business Analytics and Data Science-Center, University of Graz, Universitätsstraße 15, Graz, 8010, Austria*
[b] *Know Center & Graz University of Technology, Sandgasse 36, Graz, 8010, Austria*

ARTICLE INFO

JEL classification:
C65
C88
Z19

Keywords:
Transcription
Local
Whisper
AI
Machine learning
Qualitative research
Interview transcription
Qualitative data analysis

ABSTRACT

Research in behavioral and experimental finance becomes more multifaceted and the analysis of data from speech interactions more important. This raises the need for technical support for researchers using qualitative data generated from speech interactions. aTrain serves this need and is an open-source, offline transcription tool with a graphical interface for audio data in multiple languages. It requires no programming skills, runs on most computers, operates without internet, and ensures data is not uploaded to external servers. aTrain combines OpenAI's Whisper transcription models with speaker recognition and provides output that integrates with MAXQDA and ATLAS.ti. Available on the Microsoft Store for easy installation, its source code is also accessible on GitHub. aTrain, designed for speed on local computers, transcribes audio files at 2-3 times the audio duration on mobile CPUs using the highest-accuracy Whisper transcription models. With an entry-level graphics card, this speed improves to 30% of the audio duration.

1. Introduction

> "Transcribing interviews is time-consuming and potentially costly work. It can be facilitated by using a transcribing machine that has a foot pedal and earphones."
>
> (Seidman, 2013, p. 118)

While empirical research in behavioral and experimental finance can be characterized as a predominantly quantitative, studies based on qualitative research data, like interviews and focus groups, have also been published (see, e.g., Bhatia et al., 2020, in the journal at hand) and the journal *Qualitative Research in Financial Markets* is even dedicated to the application of qualitative methods (https://www.emeraldgrouppublishing.com/journal/qrfm). Further, method triangulation becomes more important and interviews and other qualitative data can contribute to a better understanding of the research areas of behavioral and experimental finance as part of a mixed method approach (for a systematic literature review of the application of mixed methods approaches in finance, see Dewasiri et al., 2018).

In the case of experiments in economics, the effects of communication have often been studied, e.g., as predefined messages or as free form communication in text chats (Brandts et al., 2019). However, free form communication can also take place through other channels, like audio calls or face-to-face either in person or in video chats. It has recently also been argued that with the rise of video conferencing during the COVID pandemic, the use of video chats, can help increase external validity of experiments (Bershadskyy et al., 2023).

The analysis of such free form communication requires a transcript of the spoken word (Bershadskyy et al., 2023) and the transcription of audio data has long been a considerable cost and time factor, potentially limiting the use of such qualitative data either in studies dedicated to those methods or as part of mixed method designs. However, in the ten years after the foot pedal was hailed as the state of the art for effectively transcribing interviews, the tools available have developed rapidly with the rise of artificial intelligence (AI). While the transcription of an interview of one hour requires up to six hours of manual work (Bell et al., 2018), advances in AI-based tools have sped up this process, significantly reducing the necessary transcription work to a fraction compared to manual transcription. An

* Corresponding author.
E-mail address: armin.haberl@uni-graz.at (A. Haberl).

https://doi.org/10.1016/j.jbef.2024.100891
Received 12 October 2023; Received in revised form 8 January 2024; Accepted 12 January 2024
Available online 17 January 2024

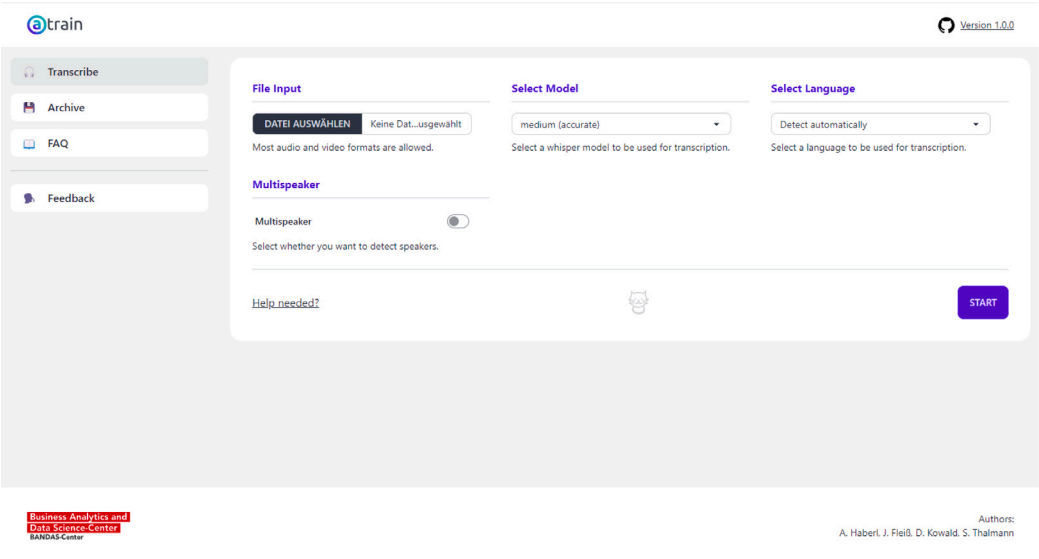

Fig. 1. aTrain user interface.

example of such an AI-based transcription model is Whisper, developed by OpenAI in 2023 (Radford et al., 2023). Whisper transcription models use a transformer architecture (Vaswani et al., 2017) to facilitate fast automated transcriptions with accuracy and robustness comparable to human transcribers (Radford et al., 2023). However, speed and accuracy of the transcription are not the only criteria relevant to qualitative researchers. When comparing the available automated transcription tools, Wollin-Giering et al. (2023) recently outlined several criteria for the comparison of automatic transcription tools: data protection, accuracy, time spent, and costs.

Although open-source transcription models such as Whisper perform generally well in all these criteria (Wollin-Giering et al., 2023), they still lack ease of use and rely on command-line interfaces. This can be a barrier for researchers without programming knowledge, especially when installed locally and operated on their own devices. They also lack output formats that integrate into common QDA software such as MAXQDA or ATLAS.ti. In addition to open-source solutions such as Whisper, there are several paid subscription tools from commercial providers such as Trint, Descript, or Sonix (see, e.g., Liyanagunawardena, 2019; Wollin-Giering et al., 2023) that provide sufficient ease of use for researchers. The problem with some of these existing tools is their cloud-native nature, which requires users to upload their interviews to external servers for transcription. Especially in the context of the EU's data protection legislation (GDPR), this practice of uploading potentially sensitive personal information to cloud services requires explicit consent from interviewees (European Parliament, Council of the European Union, 2016), which they may not be willing to give. Also, from the perspective of ethics committees approving qualitative studies, privacy preserving technologies are demanded and privacy concerns when using online transcription services for sensitive recordings have been reported in the media in the past (Mitchell, 2022). The latter concern could also arise in studies on financial markets, e.g., for interviews with whistleblowers or market participants engaging in illegal activities like antitrust violations.

To help researchers use high-quality open source tools for local interview transcription, we developed a tool providing accessible transcription of interviews: aTrain.

We developed aTrain as a free, open-source, encapsulated, self-installing, and completely offline alternative to existing solutions. The following article presents the main features and details of the technical implementation and benchmarks of the transcription time on different computer configurations. We alleviate three main issues in using free open-source automated transcription on local computers: long runtime, difficulty in setup and use, and integration in the qualitative data analysis software workflow.

To reduce runtime issues and improve on other open source solutions, we make use of the faster-whisper framework, which reduces runtime on CPUs by the factor of 4 to 5 (Klein, 2023), thus making local transcription on typical business notebooks feasible. In our benchmarks, we found that mid-range business notebooks, as are common in university settings, allow one to transcribe one hour of audio with the highest accuracy Whisper transcription models in only about 2 and a half to four hours (depending on the hardware). Additionally, aTrain also supports CUDA-enabled NVIDIA graphics cards, drastically reducing the time required to transcribe interviews.

Regarding ease of use, we offer a Microsoft Store deployment for Windows users and a simple graphical user interface Fig. 1, thus eliminating the barriers of command line setup and use of Whisper. Furthermore, we also offer an export format that allows integration in MAXQDA and ATLAS.ti to allow researchers to continue directly working in established software tools after transcription with aTrain (see Fig. 2).[1]

Finally, addressing integration in QDA software workflow, we provide an output format for syncing audio with transcript passages in the widely used software tools MAXQDA and ATLAS.ti.

2. Main features

This section provides an overview of the features that the current version of aTrain offers its users. For an in-depth description of the technical implementation of these features, refer to later chapters.

2.1. Local installation

To simplify the installation process of aTrain as much as possible, we developed a MSIX package of the application and provide it through the Microsoft Store (see e.g., Microsoft, 2021). The installation does not require administrator rights on the local machine and should even work on systems that are centrally administrated, for example, by

[1] To the best of our knowledge, one other open source tool has been developed to alleviate difficulty in setup and use: noScribe. However, the issues of long runtime on CPUs and lack of GPU support have been pointed out as a trade-off that must be made when using this solution (Wollin-Giering et al., 2023), and that we hope to alleviate.

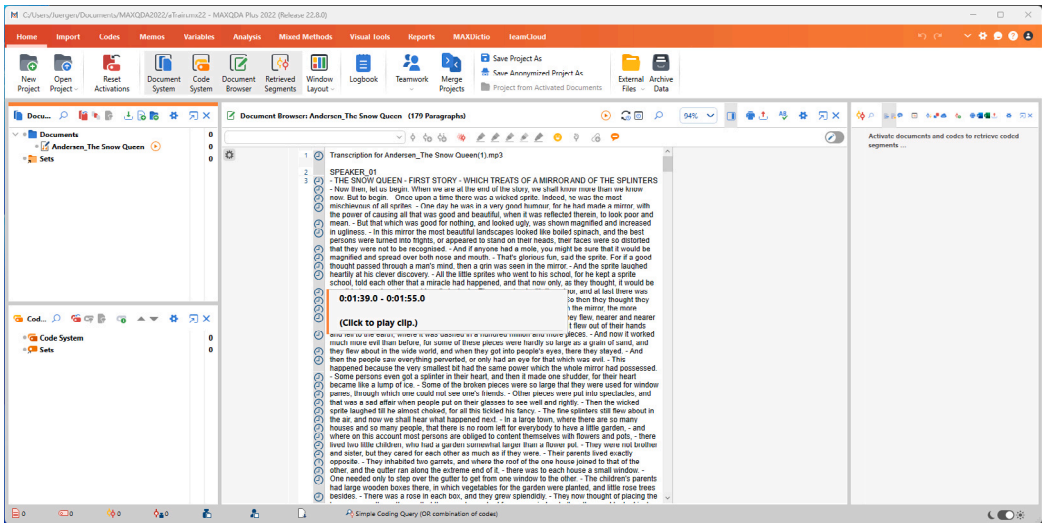

Fig. 2. Example of a transcript with synced audio in MAXQDA 2022.

university IT departments. aTrain is available through the following link to the Microsoft Store: apps.microsoft.com/store/detail/atrain/9N15Q44SZNS2.

For institutions that disabled the Microsoft Store for their users, there are also alternative installers available on the website of aTrain: business-analytics.uni-graz.at/en/research/atrain/download/.

The source code can be found in the GitHub repository under github.com/JuergenFleiss/aTrain.

2.2. Automated transcription

The central feature of aTrain is its ability to automatically transcribe audio and video files containing speech. aTrain will use its integrated Whisper transcription models to transcribe the contained speech and output the corresponding text, including time stamps. Additional settings for this transcription feature will be discussed in Section 3.

It should be noted that in qualitative research, several distinct forms of transcriptions are differentiated, with varying level of detail and the in- and exclusion of phonetic components. A systematic discussion of the different systems and how AI-based translation relates to them is presented by Wollin-Giering et al. (2023) who conclude that currently all automated systems omit phonetic information. This excludes certain types of qualitative analysis methods that focus on or include phonetic information. Instead, they provide a reproduction of the text approaching a verbatim transcription of what was said, but omitting information like pauses or filler words. However, while the system on which we built our tool, Whisper, was found to be the most accurate and detailed (Wollin-Giering et al., 2023), it is essential that the researcher reviews the transcript and compares it with the original recording. We highly recommend that researchers using aTrain consult (Wollin-Giering et al., 2023) for a systematic review and discussion of current limitations, capabilities, and advantages of automated transcription tools.

2.3. Output formats

The transcribed text will be exported from aTrain as a text file that contains timestamps for each text segment and optionally information about the corresponding speakers. The following files are provided:

- A plain text file containing the transcripts including timestamps and, if selected, speaker information.
- A plain text file containing the transcripts without timestamps and, if selected, speaker information.
- A version of the plain text file formatted for QDA software import.
- A json file containing the complete raw-transcript information, allowing users to construct their own output formats.

To ensure easy integration into existing research workflows, we also provide an output format that was designed for seamless import into the commonly used QDA software tools ATLAS.ti and MAXQDA. This then allows syncing the audio or video file to the corresponding points in the transcript, thus playing the corresponding audio or video file by clicking the corresponding text in the transcript.

2.4. Offline execution and data privacy

aTrain is designed to run completely offline without the need for an internet connection. Thus, both the Whisper transcription model and the data set operate in memory and do not leave the local machine. The same is true for the transcribed text. This means that no data is sent to any server, which ensures data privacy as part of GDPR compliance.

The version of aTrain referenced in this article and its components have been reviewed with regard to processing data exclusively offline by the authors by analyzing the source code of all components. To ensure that updates to the used components do not introduce the transfer of locally stored data to third parties, we include all components in the provided installer. This results in a relatively large installer size of around 12.8 GB. Finally, it should be noted that aTrain does not transfer the data to external servers, but that other data breaches can happen on local machines.

3. Settings

There are several settings within aTrain that allow users to finetune the transcription process. The following section gives an overview of the available options.

3.1. File input

To start a transcription with aTrain the user has to select an audio or video file from the local hard drive. aTrain accepts file inputs in most audio and video formats, since it internally converts the input file into a suitable format using the ffmpeg library (ffmpeg, 2023). A full list of supported formats can be found on the website of ffmpeg: ffmpeg.org/ffmpeg-formats.html.

Table 1
Word error rates of Whisper transcription models in various languages (Radford et al., 2023).

Model	Dutch	English	French	German	Italian	Polish	Portuguese	Spanish
Tiny	39.4	15.7	36.8	24.9	41.7	34.2	31.3	19.2
Base	28.4	11.7	26.6	17.7	31.1	22.8	21.9	12.8
Small	17.2	8.3	16.2	10.5	21.4	11.2	13.0	7.8
Medium	11.7	6.8	8.9	7.4	16.0	6.5	9.0	5.3
Large	10.2	6.3	8.9	6.6	14.3	6.6	9.2	5.4
Large-v2	9.3	6.2	7.3	5.5	13.8	5.0	6.8	4.2

3.2. Whisper transcription models

aTrain offers the option to specify the Whisper transcription model which should be used for transcription. Our software includes six different Whisper transcription models, ranging from the tiny model to the large-v2 model (Radford et al., 2023). These Whisper transcription models differ noticeably in transcription quality as shown by their respective Word Error Rates in Table 1. While larger models deliver better transcription results, they also take more time to transcribe the provided input file. A comparison of processing times on different machines is provided in Section 5.

3.3. Language

The inputted recordings can contain speech in any of the 57 languages available in the underlying Whisper transcription models used by aTrain (Radford et al., 2023). While aTrain can transcribe speech in any of these languages, the quality of the transcriptions is generally better for the languages that were predominant in the underlying training data sets (Radford et al., 2023). Users can specify the language of their recording, or let aTrain detect the language automatically based on the first 30 s of the recording. Currently, only a single language can be processed for the entire input file.[2] It is also possible to create English transcriptions for non-English source material by manually setting the language option to English. Translation to other languages is not supported.

3.4. Speaker detection

aTrain additionally offers speaker detection based on Pyannote.Audio (see Bredin et al., 2020), assigning text passages to the corresponding speakers. Users can either set the number of speakers in a recording, or let aTrain automatically detect the number of speakers. We recommend specifying the speaker count, as we found that it improves the accuracy of the clustering algorithm used for speaker detection.

3.5. Advanced settings

There are additional settings in aTrain, that are only available if a computer has a CUDA-enabled Nvidia GPU. In this case, the transcription process will run on the GPU and use a different underlying data type.[3] If users encounter problems with these settings, they can opt out of GPU usage and instead use the standard CPU settings.

[2] One workaround would be to split your audio file into segments based on language manually, using an audio or video editing tool. This would only be feasible, of course, for larger segments with the same language.

[3] aTrain uses half-precision floating-point numbers (float16) for the inference on GPUs. However, there are older GPUs that do not support computation with float16. Therefore, we added the option to use 8-bit integer values (int8) instead, which we use by default on CPUs. Rounding to int8 can lead to less accuracy but enables faster inference and less memory usage (see e.g., Nagel et al., 2021).

4. Technology and programming

4.1. Hard- and Software requirements

We tested aTrain on various computer configurations to ensure that it runs on different hardware as well as on more powerful systems with a CUDA-enabled Nvidia GPU.[4] While we expect aTrain to also work on similar and less powerful hardware, we do not have data to definitively support this claim.

aTrain can run on a CPU or on a CUDA-enabled NVIDIA GPU, the latter requiring the installation of the CUDA toolkit (NVIDIA, 2023). While a CUDA-enabled NVIDIA GPU is not required to run aTrain, it significantly improves the speed of transcriptions and speaker detection, as reported in Section 5. aTrain is available as packaged software for Windows 10 and Windows 11 and does not require any additional software dependencies. For Debian-based Linux distributions, aTrain can be installed manually from the source code, which is described in the wiki section of the aTrain GitHub Repository.[5] macOS is currently not supported but might be added in later versions of aTrain.

4.2. Software packages and architecture

The general architecture of the aTrain desktop application is based on web technologies using the Python programming language and the flask web framework (Pallets Projects, 2021). The user interface was built with HTML, CSS and JavaScript utilizing libraries such as tailwindCSS (Tailwind Labs, 2023), daisyUI (Saadeghi, 2023), alpine (Porzio, 2023) and htmx (Big Sky Software, 2023).

The main pipeline used for transcription and speaker detection includes several ML models and data processing steps. An important goal in composing this pipeline was to reduce the time needed for transcriptions and speaker detection and to limit the number of ML models included in the final MSIX installer. The final pipeline is depicted in Fig. 3.

The pipeline is initiated once the user provides a speech recording to the system, which is first converted to the WAV format using a Python binding for the ffmpeg library (Kroenig, 2019). Subsequently, the resulting WAV file is used as input for the Whisper transcription model (Radford et al., 2023) specified by the user. aTrain uses the faster-whisper implementation of Whisper, which is optimized for speed and memory efficiency, which reduces the transcription time on CPUs by a factor of 4 to 5 (Klein, 2023). The output of the Whisper transcription model contains the transcribed text segments of the speech recording, including time stamps. To enable additional speaker detection, aTrain utilizes a modified version of the Pyannote.Audio speaker detection pipeline as described in Bredin et al. (2020). An alignment function developed by Bain et al. (2023) is ultimately used to combine the outputs of the Whisper transcription and Pyannote speaker detection models into the final output.

[4] For a list of CUDA-enabled GPUs see https://developer.nvidia.com/cuda-gpus, accessed on 02 October 2023.

[5] https://github.com/JuergenFleiss/aTrain/wiki/Linux-Support

Fig. 3. Transcription pipeline with corresponding inputs, tools and outputs.

Table 2
Hardware specifications of our experiments. Please note that the computing device used in the benchmarks is highlighted in **bold**.

Machine	Year	CPU	RAM	GPU
Dell latitude 5530	2023	**Intel i5-1245U**	16 GB	CPU integrated
Lenovo thinkpad P14s	2022	**AMD Ryzen 7 PRO 6850U**	32 GB	CPU integrated
Lenovo legion Y740	2019	Intel i7-8750H	16 GB	**Nvidia RTX 2070 MaxQ 8 GB**

4.3. Source code, collaboration and license

We developed aTrain using Git as the version control system and share the source code through GitHub under github.com/JuergenFleiss/aTrain. With this, we also want to encourage collaboration with other developers. For future versions of aTrain, we would greatly appreciate the input of other developers and qualitative researchers to expand its features.

aTrain is published under an adaptation of the MIT license, where we ask users to cite this paper when using aTrain for academic or other publications.

5. Transcription duration on different whisper transcription models

To enable a comparison of aTrain against existing implementations, we want to provide crucial performance metrics. While we did not develop any ML models ourselves in this project, we did use and combine them in a novel way. Important performance indicators for automatic transcription tools include factors such as transcription accuracy and processing time (see e.g., Wollin-Giering et al., 2023). The accuracy of aTrain's transcriptions and speaker detection depends entirely on the underlying transcription and speaker detection models developed by Radford et al. (2023) and Bredin et al. (2020). Accuracy benchmarks and error rates for those ML models were already documented in the respective papers and are therefore not duplicated in this publication. In general, however, a recent comparison between various commercial and open source solutions found Whisper to be the most accurate tool available, for both a German and an English sample interview (Wollin-Giering et al., 2023).

A main performance indicator is the total processing time that results from aTrain's transcription and speaker detection pipeline. To test this processing time, we used aTrain to transcribe a conversation between Christine Lagarde and Andrea Enria at the Fifth ECB Forum on Banking Supervision 2023 published on YouTube by the European Central Bank under a Creative Commons license (https://www.youtube.com/watch?v=kd7e3OXkajY), downloaded as 320p MP4 video file. The file has a duration of exactly 22 min and represents a possible use case of aTrain in behavioral finance research, providing transcripts of publicly available video or audio files that then can be analyzed using traditional qualitative research or data-driven methods such as text mining or sentiment analysis.

This video was transcribed with all available Whisper transcription models from the faster-whisper implementation (Klein, 2023) and with speaker detection activated (Bredin et al., 2020). Additionally, we ran the transcriptions on three different computers to test the processing time on different CPUs and also on a CUDA-enabled NVIDIA GPU. The exact specifications of the hardware used during testing and benchmarking are listed in Table 2.

The processing time is recorded by aTrain using timestamps generated at the beginning and end of every transcription and displayed as the total duration after the transcription is complete.

Fig. 4 shows the processing time of each transcription relative to the length of the speech recording. In this relative processing time (RPT), a transcription is considered 'real time' when the recording length and the processing time are equal. Subsequently, faster transcriptions lead to an RPT below 1 and slower transcriptions to an RPT time above 1.

In alignment with previous expectations, the RPT increases significantly with the size of the Whisper transcription model used. Up to the small Whisper transcription model, the RPT was consistently below 1 for every computer tested. Larger Whisper transcription models resulted in RPTs ranging between 1 and 2.5 when running on CPU, while the test machine using a CUDA-enabled NVIDIA GPU (comparable to a 2023 entry-level GPU) could run transcriptions with an RPT around 0.3 even on the largest Whisper transcription models. Using standard hardware without GPU (12th gen Intel, Ryzen 7) is nevertheless feasible when running on the medium Whisper transcription model, which can provide sufficient transcription quality and speed for many use cases. For use cases without time constraints where transcription quality is especially important, we encourage users to use larger Whisper transcription models that provide the best transcription results. The use of large Whisper transcription models without a GPU can also be feasible; especially on potent multicore CPUs, where we observed an RPT of around 1.6 for the large Whisper transcription models. Note that buying an entry-level gaming notebook with a dedicated GPU to transcribe interviews with five times and more real-time speed would cost around the same as a paid automated transcription single-user license starting at around 600€ (e.g., https://app.trint.com/plans, accessed on 3 October 2023).

6. Conclusion

Research in behavioral and experimental finance becomes more multifaceted and the analysis of data from speech interactions more important. In the paper on hand, we serve this need and introduce the open-source tool aTrain, which enables the transcription (including speaker recognition) of audio recordings. The software components in aTrain were reviewed by one of the authors with a Computer Science background to ensure that it runs completely on the local computer on which it is installed and does not send any data to online servers, thus helping researchers maintain data privacy requirements arising from ethical guidelines or to comply with legal requirements such as the GDPR.

It comes with an easy-to-use graphical interface, runs on both CPUs and CUDA-enabled NVIDIA GPUs, and is available through the Microsoft Store on Windows computers. aTrain accepts most common audio and video formats as input and produces a transcript for analysis

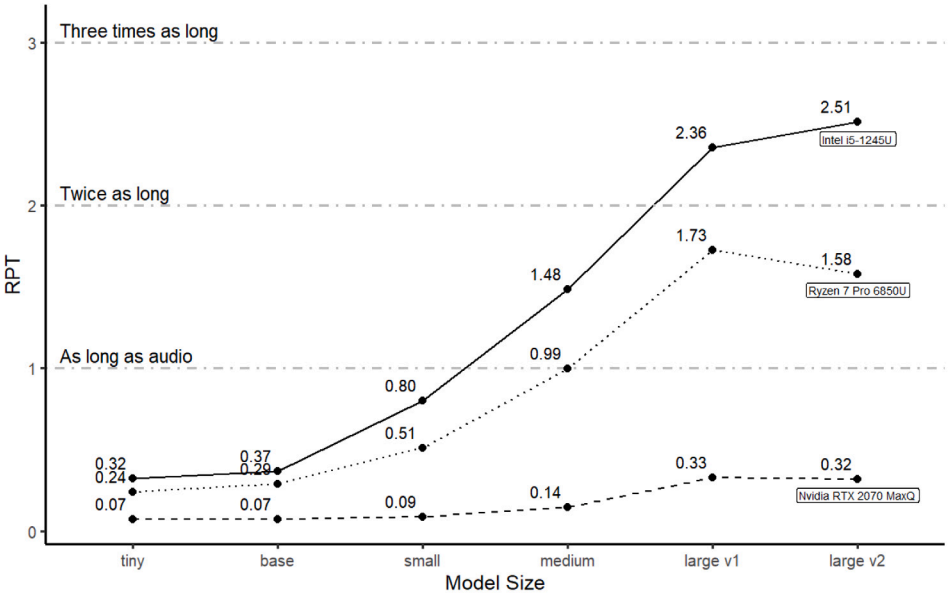

Fig. 4. Relative processing times on different hardware.

using the Whisper transcription model introduced by OpenAI, providing best-in-class accuracy among current commercial and open-source automated audio transcription solutions. It also provides an output format for import into the QDA software solutions MAXQDA and ATLAS.ti, allowing researchers to play the audio recording corresponding to text passages of the transcript with a single click.

We hope that this tool enables researchers who use audio recordings of various forms of speech interactions, either as their main data source or as a supplement to more quantitative oriented studies like laboratory experiments, to save both time and monetary resources in transcribing their audio recordings.

7. Future work

As an open-source project, aTrain will be further developed by the initial development team and independent open-source contributors. A short-term goal is the development and distribution of standalone executables and installers for Linux and MacOS. To further address possible data privacy concerns, we are also considering obtaining a certification of GDPR compliance. For this purpose, we are already in contact with external certification institutions.

CRediT authorship contribution statement

Armin Haberl: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft. **Jürgen Fleiß:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Writing – original draft. **Dominik Kowald:** Investigation, Software, Validation, Writing – review & editing. **Stefan Thalmann:** Conceptualization, Resources, Supervision, Validation, Writing – review & editing.

References

Bain, M., Huh, J., Han, T., Zisserman, A., 2023. WhisperX: Time-accurate speech transcription of long-form audio. In: INTERSPEECH 2023. URL: https://arxiv.org/abs/2303.00747.

Bell, E., Bryman, A., Harley, B., 2018. Business Research Methods. Oxford University Press, Oxford.

Bershadskyy, D., Dinges, L., Fiedler, M.-A., Al-Hamadi, A., Ostermaier, N., Weimann, J., 2023. Experimental Economics for Machine Learning-A Methodological Contribution. Working Paper Series.

Bhatia, A., Chandani, A., Chhateja, J., 2020. Robo advisory and its potential in addressing the behavioral biases of investors — A qualitative study in Indian context. J. Behav. Exp. Financ. 25, 100281. http://dx.doi.org/10.1016/j.jbef.2020.100281, URL: https://www.sciencedirect.com/science/article/pii/S2214635019302394.

Big Sky Software, 2023. HTMX: Version 1.9.3. URL: https://github.com/bigskysoftware/htmx/releases/tag/v1.9.3. (Accessed 28 September 2023).

Brandts, J., Cooper, D., Rott, C., 2019. Communication in laboratory experiments. In: Schram, A., Ulze, A. (Eds.), Handbook of Research Methods and Applications in Experimental Economics. Edward Elgar Publishing, Cheltenham, UK, pp. 401–418.

Bredin, H., Yin, R., Coria, J.M., Gelly, G., Korshunov, P., Lavechin, M., Fustes, D., Titeux, H., Bouaziz, W., Gill, M.-P., 2020. Pyannote.audio: Neural building blocks for speaker diarization. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 7124–7128. http://dx.doi.org/10.1109/ICASSP40776.2020.9052974.

Dewasiri, N.J., Weerakoon, Y.K.B., Azeez, A.A., 2018. Mixed methods in finance research: The rationale and research designs. Int. J. Qual. Methods 17 (1), 1609406918801730. http://dx.doi.org/10.1177/1609406918801730.

European Parliament, Council of the European Union, 2016. Regulation (EU) 2016/679 of the European parliament and of the council. URL: https://data.europa.eu/eli/reg/2016/679/oj. (Accessed 13 April 2023).

ffmpeg, 2023. Ffmpeg: Version 5.1.4. URL: https://github.com/FFmpeg/FFmpeg/releases/tag/n5.1.4. (Accessed 23 September 2023).

Klein, G., 2023. Faster-whipser: Version 0.7.1. URL: https://github.com/guillaumekln/faster-whisper/releases/tag/v0.7.1. (Accessed 28 September 2023).

Kroenig, K., 2019. Ffmpeg-python: Version 0.2.0. URL: https://github.com/kkroening/ffmpeg-python/releases/tag/0.2.0. (Accessed 28 September 2023).

Liyanagunawardena, T.R., 2019. Automatic transcription software: good enough for accessibility? A case study from built environment education. In: European Distance and E-Learning Network (EDEN) Conference Proceedings. European Distance and E-Learning Network, pp. 388–396.

Microsoft, 2021. What is MSIX?. URL: https://learn.microsoft.com/en-us/windows/msix/overview. (Accessed 15 September 2023).

Mitchell, C., 2022. This journalist's otter.ai scare is a reminder that cloud transcription isn't completely private. A reminder of the tradeoffs for ease and simplicity. Verge URL: https://www.theverge.com/2022/2/16/22937766/go-read-this-otter-ai-transcription-data-privacy-report. (Accessed 20 December 2012).

Nagel, M., Fournarakis, M., Amjad, R.A., Bondarenko, Y., van Baalen, M., Blankevoort, T., 2021. A white paper on neural network quantization. http://dx.doi.org/10.48550/ARXIV.2106.08295, URL: https://arxiv.org/abs/2106.08295.

NVIDIA, 2023. CUDA toolkit documentation 12.2 update 2. URL: https://docs.nvidia.com/cuda/. (Accessed 23 September 2023).

Pallets Projects, 2021. Flask: Version 2.3.3. URL: https://github.com/pallets/flask/releases/tag/2.3.3. (Accessed 28 September 2023).

Porzio, C., 2023. Alpine: Version 3.13.0. URL: https://github.com/alpinejs/alpine/releases/tag/v3.13.0. (Accessed 28 September 2023).

Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I., 2023. Robust speech recognition via large-scale weak supervision. In: International Conference on Machine Learning. PMLR, pp. 28492–28518.

Saadeghi, P., 2023. DaisyUI: Version 3.5.1. URL: https://github.com/saadeghi/daisyui/releases/tag/v3.5.1. (Accessed 28 September 2023).

Seidman, I., 2013. Interviewing As Qualitative Research: A Guide for Researchers in Education and the Social Sciences. Teachers College Press, New York and London.
Tailwind Labs, 2023. TailwindCSS: Version 3.3.3. URL: https://github.com/tailwindlabs/tailwindcss/releases/tag/v3.3.3. (Accessed 23 September 2023).
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.
Wollin-Giering, S., Hoffmann, M., Höfting, J., Ventzke, C., 2023. Automatic Transcription of Qualitative Interviews. Sociology of Science Discussion Papers, URL: https://www.static.tu.berlin/fileadmin/www/10005401/Publikationen_sos/Wollin-Giering_et_al_2023_Automatic_transcription.pdf.