BAS Web Services for Automatic Subtitle Creation and Anonymization

Florian Schiel, Thomas Kisler

Institute of Phonetics and Speech Processing, LMU Munich, Germany

schiel,kisler@phonetik.uni-muenchen.de

Abstract

In this Show&Tell contribution we will demonstrate two new public web services provided by the CLARIN centre Bavarian Archive for Speech Signals at the university of Munich. 'Subtitle' is a service that allows users to automatically create and add a subtitle track to video recordings; 'Anonymizer' can be applied to media files and their respective annotations in order to mask user-defined spoken terms in the signal as well as in the annotation. Both services are accessible via a RESTful API or a user-friendly web-interface. In the demo we will demonstrate both services independently and in combination (anonymizing subtitles) using the web interface.

Index Terms: web service, subtitle, anonymization

1. Introduction

Web services and interfaces are an elegant way to provide valuable services to users that do not necessarily possess the programming skills to implement the same functionality on their local computer system. Within the CLARIN initiative ([1]) a multitude of such web services/interfaces is provided for the humanities and other scientific fields, mostly concerned with the processing of text and speech data (see e.g. the WebLicht text processing pipeline [2]). The basic idea is that a stable service API is provided by a maintainer of an infrastructure (a university, a computing center, etc.) that can be accessed either directly (for instance from within another application), or indirectly via a user-friendly web-based interface that can be used with any web browser. It should be stressed that such a service must be stable in the sense that all further versions have to be backwards compatible, and reliable, i.e. accessible 24/7 with as little downtime as possible.

The Bavarian Archive for Speech Signals (BAS) provides such services since 2015 within the CLARIN framework. CLARIN services must fulfill certain requirements, such as stability, documentation and published metadata information[1], and are subject to bienal evaluations by the CLARIN consortium and CoreTrustSeal[2]. BAS web services are dedicated to the multi-lingual processing of speech resources, i.e. speech and multimedia corpora and their corresponding annotations; the BAS web services at the time of writing comprise automatic speech recognition (ASR), automatic speech chunking (CHUNKER, CHUNKPREP), grapheme-to-phoneme conversion (G2P), phonetic segmentation (MINNI), automatic segmentation & labeling (MAUS), text alignment, syllabification (PHO2SYL), speech synthesis (MARY), and a service that combines several web services into one processing pipeline[3].

In this Show&Tell we will demonstrate two newly published services at BAS, 'Subtitle' and 'Anonymizer' as well as their combination, using the BAS CLARIN web interface[4].

Subtitling is a convenient way to visualize transcripts or other speech annotations together with the source video. It is often used within the Digital Humanities on informant interviews such as applied in Oral History (e.g. [3]). However, subtitling is a complex task that requires programming skills, especially if the video has not yet been transcribed. In the demo we will demonstrate how a short video (30sec) can be augmented by a subtitled transcript automatically.

When publishing empirical recording materials scientists must follow certain ethical and legal requirements. Often it is necessary to 'anonymize' recordings and annotations before publication, so that certain names or phrases are not recognizable in the published data. Depending on the material and the number of words/phrases concerned, this can be a very tedious task, and many users are not capable of performing such signal masking without special training. The service 'Anonymizer' at BAS allows researchers to anonymize their media files and corresponding annotations in one processing step.

In the following we will briefly outline the principles and usage of both BAS web services.

2. Automatic Subtitle Generation

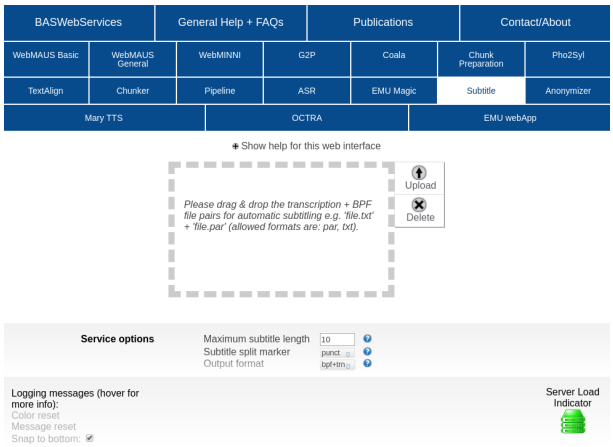

Figure 1: *BAS web interface for service 'Subtitle'.*

Fig. 1 shows a screenshot of the BAS web interface for the 'Subtitle' service[5]. To create a subtitle track one basically needs to know the time-alignment of all words in the transcript. Since this information can be obtained using other BAS services or external tools, the 'Subtitle' service does not perform its own

[1] For more details about BAS CLARIN web services refer to hdl.handle.net/11858/00-1779-0000-000C-DAAF-B.

[2] www.coretrustseal.org

[3] See clarin.phonetik.uni-muenchen.de/BASWebServices/services/help for definitions of the current web service API.

[4] hdl.handle.net/11858/00-1779-0000-0028-421B-4

[5] clarin.phonetik.uni-muenchen.de/BASWebServices/interface/Subtitle

time-alignment, but rather reads this information from the input annotation file (which must be formatted as BAS Partitur Format, BPF)[6]. The BAS 'Subtitle' service simply renders this information such that words are grouped in easily readable subtitle text blocks and outputs this rendering into file formats that can be interpreted by video players.

Consequently, when creating subtitles the user will in most cases apply the service not as a single service but rather in a processing chain, a pipeline, that first performs other services, such as automatic speech recognition or time-alignment, in preparation for the final service 'Subtitle'. There are of course many different ways to embed the 'Subtitle' service into a pipeline depending on the type and properties of the input data. For instance, if the input video is very long, it would make sense to include a service module called 'Chunker' into the initial pipeline which breaks up the large signal file into workable chunks.

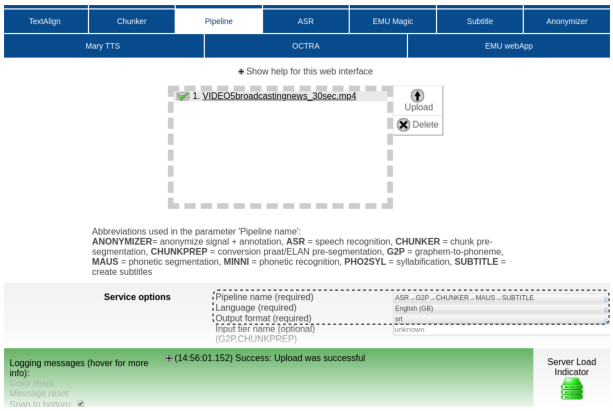

Figure 2: *BAS web interface for the pipeline service employing the 'Subtitle' service. The selected pipeline is 'ASR_G2P_CHUNKER_MAUS_SUBTITLE'; the processing language is British English; the output format is SubRip Subtitle (srt).*

For the sake of brevity, we will only demonstrate one type of pipeline employing 'Subtitle' as the last module. This pipeline creates a subtitle track, taking only the video as input. Consequently, the first service in the pipeline is automatic speech recognition ('ASR'), which provides us with a rough transcript of what is being said in the video. The next service 'G2P' transforms the text into a phonetic form performing a grapheme-to-phoneme conversion. This is followed by the 'Chunker' service which segments the video into smaller chunks, and the 'MAUS' service which provides us with the necessary word-time alignment. The web interface[7] showing the pipeline 'ASR_G2P_CHUNKER_MAUS_SUBTITLE' can be seen in Fig. 2 with the input video file already uploaded to the server.

As can be seen from Fig. 1 there are two options[8] that influence how words are grouped and displayed as subtitles: the 'Subtitle split marker' option defines where the input transcript is split into subtitle blocks (alternatives are: punctuation, newline, or a customized tag); the 'Maximum subtitle length' defines the maximum number of words per subtitle, i.e. if no split marker can be found, the subtitle is truncated to this number.

[6]See www.bas.uni-muenchen.de/forschung/Bas/BasFormatseng.html for details about the BPF.

[7]clarin.phonetik.uni-muenchen.de/BASWebServices/interface/Pipeline

[8]In the pipeline web interface these options can be found under 'Expert options'.

3. Automatic Anonymization

The second web service demonstrated in this Show&Tell is 'Anonymizer', a tool that is very similar to 'Subtitle' (Section 2) in that it requires the time-alignment of words as a necessary input. In isolation, 'Anonymizer' will read a media file and a corresponding BPF annotation file as input and produce a media file and an annotation file as output[9]. Additionally, the user must provide the service with a simple list of terms to be anonymized ('aTerms' in the following). aTerms may contain white space and an unlimited number of words, which should be listed on separate lines in a UTF-8-encoded text file. The 'Anonymizer' then searches for literal matches of listed aTerms in the orthographic layer of the input BPF file. Each time a match is found, the service replaces the labels in all annotation layers assigned to an aTerm by a predefined string, and all phonetic labels (e.g. phonemes, syllables) by another predefined label; both label strings can be set by the user via service options. In parallel, the service calculates the time location of the aTerm within the media file and masks the speech signal with either brown noise or a low sine wave tone (400Hz, -6dB).

As with the 'Subtitle' service, the 'Anonymizer' is most likely applied within a pipeline. We therefore demonstrate the usage of this service in a pipeline that anonymizes the subtitle track of Section 2: 'ASR_G2P_CHUNKER_MAUS_ANONYMIZER_SUBTITLE'. Fig. 3 shows the pipeline web interface with the selected pipeline and the selected options for the 'Anonymizer' module.

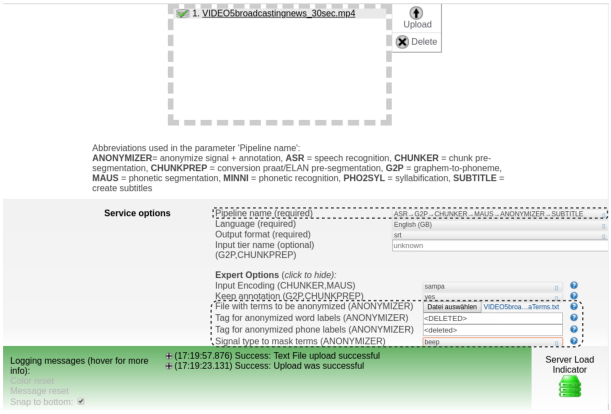

Figure 3: *BAS web interface for the pipeline service using the 'Anonymizer' service together with selected options.*

4. References

[1] E. Hinrichs and S. Krauwer, "The CLARIN Research Infrastructure: Resources and Tools for e-Humanities Scholars," *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, pp. 1525–1531, May 2014. [Online]. Available: http://dspace.library.uu.nl/handle/1874/307981

[2] E. W. Hinrichs, M. Hinrichs, and T. Zastrow, "Weblicht: Web-based lrt services for german," in *Proceedings of the ACL 2010 System Demonstrations*, 2010, pp. 25–29. [Online]. Available: http://www.aclweb.org/anthology/P10-4005

[3] P. Thompson, *The voice of the past: Oral history*. Oxford university press, 2017.

[9]The user can select from four different output formats.