https://www.doi.org/10.21272/Ftrk.2024.16(2)-25

THE EDUCATIONAL AUDIO LIBRARY FOR EFL LEARNERS

Anokhina Tetiana,
Kyiv National Linguistic University, Ukraine
ORCID ID 0000-0002-8859-5568
Corresponding author: tetiana.anokhina@knlu.edu.ua

Kobyakova Iryna,
Sumy State University, Ukraine
ORCID ID 0000-0002-9505-2502

Schvachko Svitlana,
Sumy State University, Ukraine
ORCID ID 0000-0002-2119-1884

Abstract. *In order to reach language proficiency, the English language learners are to enhance their English mastery in class using audio and video training materials. We aim at training students to achieve high level or near-native fluency. To perform this task we are training listening and comprehension skills using modern libraries of contemporary speech flows. With the rise of digital media, we can compile the self-made corpus of audio and video materials. Thus, we are to find the needful content that can be tracked from academic or professional issues available online. This task can be performed manually or semi-automatically, with the help of available tools and libraries. We find it useful to be able to compile the library of media libraries for A1, B1, C1, C2, C3 and Proficiency levels. The bank can serve to different levels based on the same resource corpus.*

Keywords: *media library, audio and video files, media libraries, corpus, A1, B1, C1, C2, C3, Proficiency level.*

Received: *11 July 2024*
Revised: *15 July 2024*
Accepted: *26 July 2024*
How to cite: Anokhina T., Kobyakova I., Schvachko Sv. (2024). The Educational Audio Library for EFL Learners. *Philological Treatises, 16*(2). https://www.doi.org/10.21272/Ftrk.2024.16(2)-25

СТВОРЕННЯ АУДІОБІБЛІОТЕКИ ДЛЯ НАВЧАННЯ АНГЛІЙСЬКОЇ МОВИ ЯК ІНОЗЕМНОЇ

Анохіна Тетяна,
Київський національний лінгвістичний університет, Україна
ORCID ID 0000-0002-8859-5568
Автор, відповідальний за листування: tetiana.anokhina@knlu.edu.ua

Кобякова Ірина,
Сумський державний університет, Україна
ORCID ID 0000-0002-9505-2502

Швачко Світлана,
Сумський державний університет, Україна
ORCID ID 0000-0002-2119-1884

Анотація. *На сучасному етапі викладання англійської як іноземної учні постійно мають вдосконалювати свої знання за допомогою аудіо- та відеоматеріалів, щоб досягти певного рівня володіння мовою. Викладачі намагаються навчити учнів досягти просунутого рівня володіння мовою та рівня, близького до володіння рідною мовою. Для виконання цих задач маємо натреновувати навички з аудіювання, використовуючи сучасні аудіо-бібліотеки сучасного живого мовлення. З появою цифрових медіа ми можемо власноруч укладати корпус аудіо- та відеоматеріалів. Тобто, ми можемо знайти необхідний вміст, який можна відстежити з академічних або професійних джерел, доступних в Інтернеті. Це завдання по створенню бібліотеки можна виконати самостійно або за допомогою програмних способів, напівавтоматично, за допомогою доступних інструментів та бібліотек. Ми вважаємо корисною можливість скомпілювати аудіобібліотеку для рівнів A1, B1, C1, C2, C3 та рівня вільне володіння (Proficiency). Корпус може функціонувати як аудіобанк різних рівнів на основі базових аутентичних та згенерованих автоматично аудіоресурсів.*

Ключові слова: *аудіобанк, аудіо- та відеофайли, аудіоресурси, корпус, A1, B1, C1, C2, C3, вільне володіння (Proficiency).*

Отримано: *11 липня 2024 р.*
Отримано після доопрацювання: *15 липня 2024 р.*
Затверджено: *26 липня 2024 р.*
Як цитувати: Анохіна Т., Кобякова І., Швачко Св. (2024). Створення аудіобібліотеки для навчання англійської мови як іноземної. *Філологічні трактати*, *16*(2). https://www.doi.org/10.21272/Ftrk.2024.16(2)-25

Introduction

Today's life is demanding and we are to use less time for better results. As we compile libraries of the modern media, we use non-adapted video and audio fragments for Proficiency level. But it is important to differentiate levels of listening and adapting to A1 and B1 learners, with further layering to C1, C2 and C3 levels. In order to compile the differentiated levels, we can use automatic speech tools such as Elevel Labs or similar ones that have paid subscription and can provide the perfect results.

This paper *(i)* proposes the use of audio splitting to prepare content of media libraries with extended pauses for A1, B1, B2, C1, C2 users and Proficiency, extracted using Dlib [6, 7] and *(ii)* compares different ways of audio programmes.

The Common European Framework of Reference for Languages (CEFR) categorises English proficiency into several levels, each characterised by distinct speech patterns, particularly the length of pauses during speech (*International Language Standards*, n.d.). We are compiling the library for training future interpreters based on the levels with a focus on the length of pauses.

At this initial stage (A1 Beginner: CEFR Level A1) speakers exhibit very limited knowledge of the English language. Their speech is notably slow, characterised by very long pauses as they search for the correct words. This level reflects a vocabulary of around 700 words and typically requires approximately 100 hours of study to achieve.

Speakers at the A2 level **(**A2 Pre-Intermediate: CEFR Level A2**)** can engage in everyday small talk and express simple opinions. While their conversations are still marked

by long pauses, these are shorter than those observed at the A1 level. They are beginning to explore basic past and future tenses but often rely on native speakers to maintain the flow of conversation. YouTube videos can be a good source to start with.

Achieving B1 proficiency **(**B1 Intermediate: CEFR Level B1) indicates a greater degree of confidence in English. Speakers can manage everyday interactions, such as shopping or dining out, with less difficulty. Although they may still struggle with unfamiliar topics, their speech is characterised by moderate pauses, demonstrating improved fluency compared to the A2 level.

At the B2 level (B2 Upper-Intermediate: CEFR Level B2) speakers demonstrate basic English fluency. They feel comfortable in English-speaking environments, such as workplaces and educational institutions. Their speech contains shorter pauses, allowing for more natural and fluid conversations. They can produce complex sentences and engage in discussions more effectively.

Speakers at the C1 level **(**C1 Advanced: CEFR Level C1) exhibit a high degree of proficiency. They can handle complex discussions and perform well in academic settings. Their speech is near-fluent, characterised by minimal pauses, indicating a strong command of the language.

The C2 level represents near-native fluency (C2 Mastery: CEFR Level C2)**.** Speakers at this level have full confidence and control over the English language. They can discuss any topic with nuanced expression and coherent delivery, with very minimal pauses. This proficiency level ensures that speakers can use English comfortably in both academic and professional settings, almost on par with native speakers.

These descriptions provide a clear overview of the progression in speech fluency and the reduction of pauses as English proficiency improves across the CEFR levels.

In language training, especially for interpreters, it is often necessary to manipulate audio files to include pauses or split them into smaller segments. This can help trainees practise their listening and speaking skills more effectively. In this article, we will explore how to split audio files in paid and free programmes (being user friendly) and add pauses using Python (with developer's possibilities). We will also discuss how to download and process these files for training purposes.

In order to get level down, we are making txt with longer pauses which can help students to train their translation skills far better in their consecutive translation practice. Understanding proficiency level audio and video materials, students show mastery in English. Understanding subjects with nuanced expression, they can coherently deliver the information into their native tongue.

Aiming to understand spoken English without barriers, students should navigate daily life dialogues, academia observation at conferences, and other places with native speakers with ease. The current possibilities to use TTS generated speech are making better possibilities for study as a teacher can run the content more effectively.

Here we must stress on the need of the media libraries that are seen as a set of files, in other words "a curated collection of audio, video, and other multimedia files". In order to enhance language proficiency, the library should provide a rich dataset with recordings of fluent speakers, educational videos, and other relevant content.

To start with any dataset we must load and save audio files in various formats. The place to save will be our library. To enhance understanding of audio content, we might generate visualisations like waveforms, spectrograms, and pitch contours. In order to generate different levels – A1, B1, C1, C2, C3 and Proficiency levels – for our library it is possible to split audio or video files and add pauses after utterings, phrases or sentences.

1.1 User's friendly programmes for splitting audio and adding longer pauses

Educators and users might learn how to do it using existing tools or subscribing to paid resources for adding pauses to audio files, such as Adobe Audition (*Adobe: Creative*,

Marketing and Document Management Solutions, n.d.) to add pauses, cut, trim, and manipulate audio files in various ways, with the advanced features for audio restoration, multi-track editing, and more.

Auphonic is a web-based audio post-production tool that can process and improve the quality of audio files with features like automatic levelling, noise reduction, adding pauses manually (*Auphonic*, n.d.). Pro Tools is the professional digital audio workstation (DAW) used widely in the music and film industry for editing, mixing, and adding pauses to audio files (*XLN Audio*, n.d.). Reaper is a cost-effective digital audio workstation (DAW here and later in the article) that offers a wide range of features for audio editing, including the ability to add pauses, cut, and manipulate audio tracks.

Sound Forge Pro is an advanced audio waveform editor designed for professionals. It includes tools for precise audio editing, including the insertion of pauses and other manipulations ((GmbH, n.d.).Audionamix specialises in advanced audio separation technology. While not directly for adding pauses, its tools can help isolate and manipulate specific parts of an audio file (*Audionamix – Artificial Intelligence (AI) Music Tool*, n.d.). Logic Pro X is a comprehensive DAW for Mac users that offers robust audio editing capabilities, including adding pauses and other detailed manipulations (Apple, 2013).

Alternatively, we can use free tools. These can be tested. The top one that works perfectly well with no restrictions at the time is Audacity which is a free, open-source audio editor to record, edit, and add pauses to audio files of various file formats (*Audacity* ®, n.d.).

Ocenaudio is a free, easy-to-use audio editor that provides a straightforward interface for editing and adding pauses to audio files. It supports real-time effects and can handle large files efficiently (*Ocenaudio*, n.d.).

TTS maker is an online powerful online tool with certain limitations in characters and styles (Fig.1.). We can fix the pauses duration and generate voiced texts automatically (*Free Text to Speech Online*, n.d.).

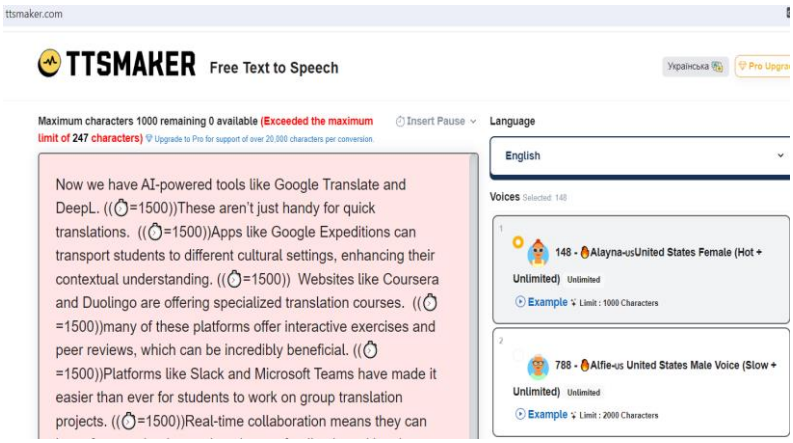

Fig. 1.1. Inserting pauses with certain limit in TTS maker

WavePad has a free version of the WavePad Audio Editing Software available for non-commercial use. It offers various editing tools, including the ability to add pauses, cut, trim, and apply effects to audio files (*Audio Editing Software. Sound, Music, Voice & MP3 Editor. Best Audio Editor for 2024.*, n.d.).

Wavosaur is a free audio editor that is lightweight and supports VST plugins. It allows you to add pauses, cut, copy, and paste audio segments, and apply various effects (*Wavosaur Free Audio Editor with VST and ASIO Support*, n.d.).

AudioMass is a free, web-based audio editor that doesn't require any downloads. You can use it directly in your browser to edit audio files, including adding pauses, trimming, and applying effects (*AudioMass*, n.d.).

Soundtrap is an online, collaborative DAW that offers a free tier. It allows you to edit audio, add pauses, and work on projects with other users in real-time (*Soundtrap - Make Music Online*, n.d.).

Articulate 360 does offers a range of TTS voices in free trial, including some with Indian accents. We might adore a bit exaggerated but natural voices produced by AI. This could be due to the technology behind the voice synthesis or the way certain linguistic nuances are handled (*Highly Customizable Course Authoring*, 2023).

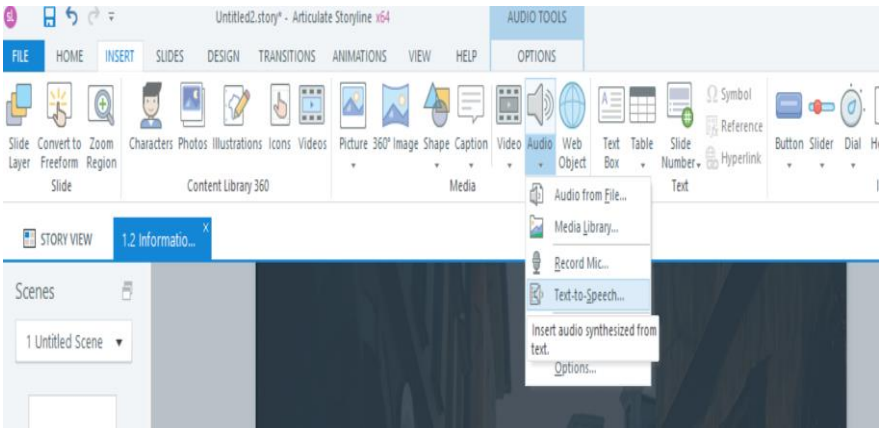

Fig. 1.2. Sending Text to Speech in Storyline by Articulate 360

TwistedWave Online is a free online audio editor that lets you edit audio files directly in your browser. You can add pauses, cut, and apply basic effects. The free version has some limitations on file length and format (*TwistedWave Online Audio Editor*, n.d.).

These tools provide a range of capabilities for editing audio files, and many of them offer enough features to meet the needs of most users.

1.2 Developer's solutions for splitting audio and adding longer pauses

If a developer is not satisfied with this functionality, one can create the other with a set of functions by integrating Librosa with media libraries to create a self-made library for audio analysis.

Librosa is a powerful Python library widely used for analysing and processing audio files. Originally designed for music information retrieval, it has found applications in various fields, including speech and language processing. Librosa can extract features from audio files of C2 speakers, analysing aspects like pronunciation, intonation, and speech rate. This helps in identifying the distinguishing characteristics of high-level proficiency. It can index media library content based on extracted features. For instance, segments showcasing exemplary use of idiomatic expressions or complex grammatical structures can be tagged and easily retrieved for educational purposes.

It can train machine learning models on annotated audio files to develop applications such as proficiency assessment tools, language learning apps, and automated transcription services.

Librosa's feature extraction capabilities are integral to these processes. It is possible to create visualisations of audio features to provide feedback to language learners. Visual tools can highlight areas for improvement, such as intonation patterns or pacing, facilitating a more targeted learning experience. To accomplish these tasks,

Librosa can be used for audio analysis and manipulation (*Librosa — Librosa 0.10.2 Documentation*, n.d.). Pydub can serve for audio file handling and manipulation

(jiaaro, n.d.). To compile the educational library the Requests also can serve for downloading audio files from the internet (Mitra, n.d.).

2. 1 Developing Educational Corpus of audio files

In order to compile the library and to generate different levels audio files – A1, B1, C1, C2, C3 and Proficiency levels we are installing the Libraries for Splitting Audio and Adding Pauses in Python.The pause durations list is now defined as [500, 1000, 1500, 2000, 2500], with 500 ms pauses for A1 (the shortest) and 2500 ms pauses for C3 (the longest). To split an audio file into smaller segments, we can use Pydub (Fig. 2.1.1).

```python
from pydub import AudioSegment

def split_audio(file_path, segment_length_ms):
    # Load the audio file
    audio = AudioSegment.from_file(file_path)

    # Calculate the number of segments
    num_segments = len(audio) // segment_length_ms

    # Split the audio into segments
    segments = [audio[i*segment_length_ms:(i+1)*segment_length_ms] for i in range(num_segm

    return segments

# Example usage
file_path = 'your_audio_file.mp3'
segment_length_ms = 10000  # Split into 10-second segments
segments = split_audio(file_path, segment_length_ms)

# Save the segments to files
for i, segment in enumerate(segments):
    segment.export(f'segment_{i}.mp3', format='mp3')
```

Fig. 2.1.1 Code fragment for splitting pauses in pydub

Alternatively we can split it in Audacity with no command line (Fig. 2.1.2).

Fig. 2.1.2 Code fragment for splitting pauses in Audacity

To add pauses to an audio file, we can create a silent segment and concatenate it with the original audio (Fig. 2.1.3).

```python
def add_pause(audio, pause_length_ms):
    # Create a silent segment
    silence = AudioSegment.silent(duration=pause_length_ms)

    # Add the silent segment to the beginning and end of the audio
    audio_with_pause = silence + audio + silence

    return audio_with_pause

# Example usage
pause_length_ms = 2000  # 2-second pause
audio_with_pause = add_pause(segments[0], pause_length_ms)
audio_with_pause.export('audio_with_pause.mp3', format='mp3')
```

Fig. 2.1.3 Code fragment for adding longer pauses

To download audio files from the internet and process them to add longer pauses, we can use the requests library (Fig. 2.1.4).

```python
import requests

def download_audio(url, file_path):
    response = requests.get(url)
    with open(file_path, 'wb') as file:
        file.write(response.content)

# Example usage
audio_url = 'https://example.com/audio_file.mp3'
download_path = 'downloaded_audio.mp3'
download_audio(audio_url, download_path)

# Load the downloaded audio
downloaded_audio = AudioSegment.from_file(download_path)

# Add pauses and save the new file
audio_with_pause = add_pause(downloaded_audio, pause_length_ms)
audio_with_pause.export('downloaded_audio_with_pause.mp3', format='mp3')
```

Fig. 2.1.4 Code fragment for downloading files with longer pauses

Conclusions

The combination of Librosa, audio files programmes with built-in TTS and pre-paid programmes like Articulate 360 with Indian accents and others offer a powerful approach to analysing and leveraging A1, B1, B2, C2 and proficiency content. By extracting and analysing audio features, annotating and indexing content, and applying machine learning techniques, researchers and educators can gain deeper insights into language proficiency and develop innovative tools for language learning and assessment. As digital technologies continues to evolve, the integration of these technologies will play a crucial role in advancing our understanding and teaching of high-level language skills.

By using TTS programmes and modern abilities to run open source we are building more relevant libraries and can create the library of our own choice and educational perspective.

REFERENCES

Adobe: Creative, marketing and document management solutions. (n.d.). Retrieved July 22, 2024, from https://www.adobe.com

Allsoft Ukraine. (n.d.). Retrieved July 23, 2024, from https://allsoft.ua/p736865687-sound-forge-elektronnaya.html

Apple. (2013, July 16). *Logic pro*. App Store. https://apps.apple.com/ua/app/logic-pro/id634148309?mt=12

Audacity ®. (n.d.). Free Audio Editor, Recorder, Music Making and More! Retrieved July 23, 2024, from https://www.audacityteam.org/

Audio editing software. Sound, music, voice & MP3 editor. Best audio editor for 2024. (n.d.). WavePad Audio Editor Free. Retrieved July 23, 2024, from https://www.nch.com.au/wavepad/index.html

AudioMass. (n.d.). Audio Editor. Retrieved July 23, 2024, from https://audiomass.co/

Audionamix - Artificial intelligence (AI) music tool. (n.d.). TuneTech.AI. Retrieved July 23, 2024, from https://www.tunetech.ai/music-ai/audionamix

Auphonic. (n.d.). Retrieved July 22, 2024, from https://auphonic.com/ Action, S. (2019). Grey worm being adorable for 5 minutes straight [Video]. In *YouTube*. https://youtu.be/WiQjKPPZdLA

Auphonic. (n.d.). Retrieved July 22, 2024, from https://auphonic.com/

Benjamine, C. (2021, February 11). *English language levels explained (and test your level!)*. Preply.Com. https://preply.com/en/blog/english-language-levels/

Commands. (n.d.). Poetry - Python Dependency Management and Packaging Made Easy. Retrieved July 11, 2024, from https://python-poetry.org/docs/cli/

Crag. (2016). Yoda teaching Skywalker Knowing [Video]. In *YouTube*. https://youtu.be/EJz2FXkZb7g

DutchHPfan1992. (2011). Harry Potter and the Chamber of Secrets - Dobby at the dursley's (HD) [Video]. In *YouTube*. https://youtu.be/Mjz0OwQmfWo

Free text to speech online. (n.d.). TTSMaker. Retrieved August 12, 2024, from https://ttsmaker.com/

Highly customizable course authoring. (2023, June 19). Articulate. https://www.articulate.com/360/storyline/

Hill, J. (2014). Angela Merkel speaking English to British Parliament [Video]. In *YouTube*. https://youtu.be/cGZWR5S1lCo

International language standards. (n.d.). Cambridge Assessment English. Retrieved July 23, 2024, from https://www.cambridgeenglish.org/exams-and-tests/cefr/

jiaaro. (n.d.). *GitHub - Jiaaro/pydub: Manipulate audio with a simple and easy high level interface*. GitHub. Retrieved July 23, 2024, from https://github.com/jiaaro/pydub

Josh H McDermott, "The cocktail party problem," *Current Biology*, vol. 19, no. 22, pp. R1024–R1027, 2009.

librosa — librosa 0.10.2 documentation. (n.d.). Retrieved July 23, 2024, from https://librosa.org/doc/latest/index.html

Live, J. K. (2018). Sofía Vergara Reveals What She Does When Husband is Away [Video]. In *YouTube*. https://youtu.be/dUO4MwwoZvM

Observations. (2018). I am groot [Video]. In *YouTube* https://youtu.be/RMdIx228dBw

Ocenadio. (n.d.). *ocenaudio*. Retrieved July 23, 2024, from https://www.ocenaudio.com/

Soundtrap - Make music online. (n.d.). Soundtrap. Retrieved July 23, 2024, from https://www.soundtrap.com/musicmakers

TwistedWave Online Audio Editor. (n.d.). Retrieved July 23, 2024, from https://twistedwave.com/online

Wavosaur free audio editor with VST and ASIO support. (n.d.). Retrieved July 23, 2024, from https://www.wavosaur.com/

XLN audio. (n.d.). XLN Audio. Retrieved July 22, 2024, from https://www.xlnaudio.com/?gad_source=1&gclid=CjwKCAjwhvi0BhA4EiwAX25uj1lbXkbQp9ksgcT89yptoa3mM87Gf-WQ0-7jQw_lj66YIFYYKWmhOhoCiWUQAvD_BwE