

SD09: Speech to Text (STT) API

1. INTRODUCTION

2. PURPOSE

In this document, we explained the public API for Bangla Speech to Text. The consumer of this API will have to follow this document to integrate it into his/her work.

3. SPECH TO TEXT (STT) API

3.1. API URL

Purpose	URL	Method
Non-streaming single file	https://stt.bangla.gov.bd/file	Post
Non-streaming batch performance	https://stt.bangla.gov.bd/batch	Post
Streaming	https://stt.bangla.gov.bd/socket.io/?EIO=4&transport=polling	Web-socket

3.2. NON-STREAMING SINGLE FILE REQUEST (FORM-DATA OBJECT)

Attribute	Type	Description	Permitted Values
model	String	Select from multiple model	crdnn, transformer (currently only 'crdnn' is available) [Default 'crdnn']
use_lm	Boolean	Whether to use external LM	True, False [default True]
file	File-part	Audio file to transcribe	Any 'wav', 'flac' or 'mp3' audio file

Default Value: 2.0

P.S. Version 1.0 is deprecated

Request Example:

```
{
  "model": "crdnn",
  "use_lm": "false",
  "file": FILE_PART,
}
```

3.3. NON-STREAMING RESPONSE (JSON OBJECT)

The response is returned in JSON format. The JSON attributes are listed below. The response will follow the standard HTTP response status code.

Attribute	Type	Description
model	String	Model name used for the transcription
lm	String	If used then LM model name otherwise does not contain this field
message	String	Any message sent from the server. If there is no message, it does not contain this field.
output	String	The transcription of the input audio file.

Response Example:

```
{
  "model": "crdnn",
  "output": "আমি বাংলায় গান গাই",
}
```

3.4 NON-STREAMING BATCH PERFORMANCE REQUEST (FORM-DATA OBJECT)

Attribute	Type	Description	Permitted Values
model	String	Select from multiple model	crdnn, transformer (currently only 'crdnn' is available) [Default 'crdnn']
use_lm	Boolean	Whether to use external LM	True, False [default True]
file	File-part	Audio file to transcribe	Any 'wav', 'flac' or 'mp3' audio file
text_file	File-part	Text file with the ground truth transcript of the speech.	Any '.txt' file.

3.5 NON-STREAMING BATCH PERFORMANCE RESPONSE (JSON OBJECT)

Attribute	Type	Description
model	String	Model name used for the transcription
audio_file_name	String	Name of the audio file that was provided
text_file_name	String	Name of the text file that was provided
lm	String	If used then LM model name otherwise does not contain this field
message	String	Any message sent from the server. If there is no message, it does not contain this field.
ground_text	String	Ground truth transcription
output	String	The transcription of the input audio file.
performance	JSON object	Performance of the model in terms of CER and WER

Response Example:

```
{  
  "audio_file_name": "t1_br_65kbps_sr_22050hz.mp3",  
  "ground_text": "হ্যালো সবাই রক্ত দিন জীবন বাঁচান রক্ত দিলে ক্ষতি নাই আমরা দলে দলে রক্ত দিতে  
যাই",  
  "model": "crdnn",  
  "output": "এতো সবাই রক্ত দিন জীবন বাঁচান রক্ত দূরে ক্ষতি নাই অন্য দলে দলে রক্ত দিতে যাই",  
  "performance": {  
    "cer": 13.924050632911392,  
    "wer": 18.75  
  },  
  "text_file_name": "t1.txt"  
}
```

3.6. STREAMING API (USING WEB-SOCKET)

Client end

Socketurl : <https://stt.bangla.gov.bd/socket.io/?EIO=4&transport=polling> (initial request)

Initial Response:

```
0{"sid":"dmk1WeeTkjXkMwciAABI","upgrades":["websocket"],"pingTimeout":20000,"pingInterval":2500  
0}
```

Final request:

```
[  
  "sendAudio",  
  {  
    "_placeholder": true,  
    "num": 0  
  }  
]
```

Binary Message

Final response:

```
[  
  "sendTranscript",  
  {  
    "msg": "আমার কথা থেকে লেখা",  
    "success": true  
  }  
]
```

Server end

Events from server side:

Event listener	Function //description	<u>(event,response) to client</u>
<u>connectListener</u>	<u>onConnected() // connects with client.</u>	<u>“client_id”, “client connected to server”</u>
<u>disconnectListener</u>	<u>onDisconnected() // disconnects with client</u>	<u>“disconnect”, “disconnected”</u>
<u>EventListener(‘chat’)</u>	<u>onByteReceived() //process audio byte and returns response</u>	<u>“stream”,transcript “disconnect now”, transcript</u>