



HAL
open science

Automated Speech Processing of Bengali using SPPAS software

Moumita Pakrashi, Shakuntala Mahanta, Brigitte Bigi

► **To cite this version:**

Moumita Pakrashi, Shakuntala Mahanta, Brigitte Bigi. Automated Speech Processing of Bengali using SPPAS software. iNNterpret - Representation of speech, articulatory dynamics, prosody and language in layers. Special Session, S2 at ICON 2021: 18th International Conference on Natural Language Processing, NLP Association India, Dec 2021, [in-line due to the pandemic], India. hal-03943762

HAL Id: hal-03943762

<https://hal.science/hal-03943762>

Submitted on 17 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automated Speech Processing of Bengali using SPPAS software

Brigette Bigi, Shakuntala Mahanta, Moumita Pakrashi

What is SPPAS?

- **SPPAS: the automatic annotation and analysis of speech**, is a tool to automatically produce annotations including utterance, word, syllable and phoneme segmentations from a speech recording and its orthographic transcription.
- <http://www.sppas.org/>
- Resulting alignments are a set of annotated files in either:
 - .xra, the XML native file of SPPAS
 - .TextGrid, the native file format of the Praat software
 - .eaf, the native file format of ELAN
 - .CSV, ...

Need of SPPAS in analysing Bengali prosody

- Scope for improving the naturalness of synthetic speech generated from present-day Bengali TTS systems.
- Intonation along with timing properties of speech prosody exercises the most prominent influence on enhancing such naturalness.
- To avoid time-consuming, manual analysis of large speech corpus, automatic analysis and annotation is essential.
- Addition of minimum linguistic resources to add a new language.

Why SPPAS?

- Free and open source software maintained and developed regularly, especially designed for linguists.
- Multi-platform software that can be used in Windows, Mac, Linux OS platforms.
- Analysis/annotation output files compatible by conversion to most speech analysis software like Praat, Elan, Annotation Pro.
- Momel-INTSINT algorithm plugin incorporated for prosodic analysis.

Primary functions

- *Annotate* - Produce automatic annotations from a recorded speech sound and its orthographic transcription and/or from a video.
- *Analyze* - Analysis of any annotated data: estimate statistical distributions, make requests, manage files, visualize annotations.
- *Convert* - SPPAS offers a file converter from/to a wide range of formats: xra, TextGrid, eaf, trs...

Resources for addition of Bengali

The speech dataset used for training the models has been obtained from the Open Speech and Language resources created by Google that is freely available.

- Acoustic model

HMM based acoustic model has been created by Brigitte using an HTK toolkit, and implemented in the SPPAS latest version 4.1.

Trained with a set of 6 manually time-aligned files (about 18 seconds of speech) and 1,300 orthographically transcribed files (36 minutes of speech).

- Pronunciation Dictionary

Pronunciation dictionary and lexicon downloaded from Google database resources. The phonemes were then converted to SAMPA and the file format to HTK-ASCII by Brigitte. Numerals had to be transcribed in Bengali for adding them to the lexicon.

- Manual alignment

Manual alignment of some amount of speech data required to evaluate the success rate of the automatic alignment.

- Syllabification

Finally syllabification rules needs to be added in order to enable automatic syllabification for Bengali utterances.

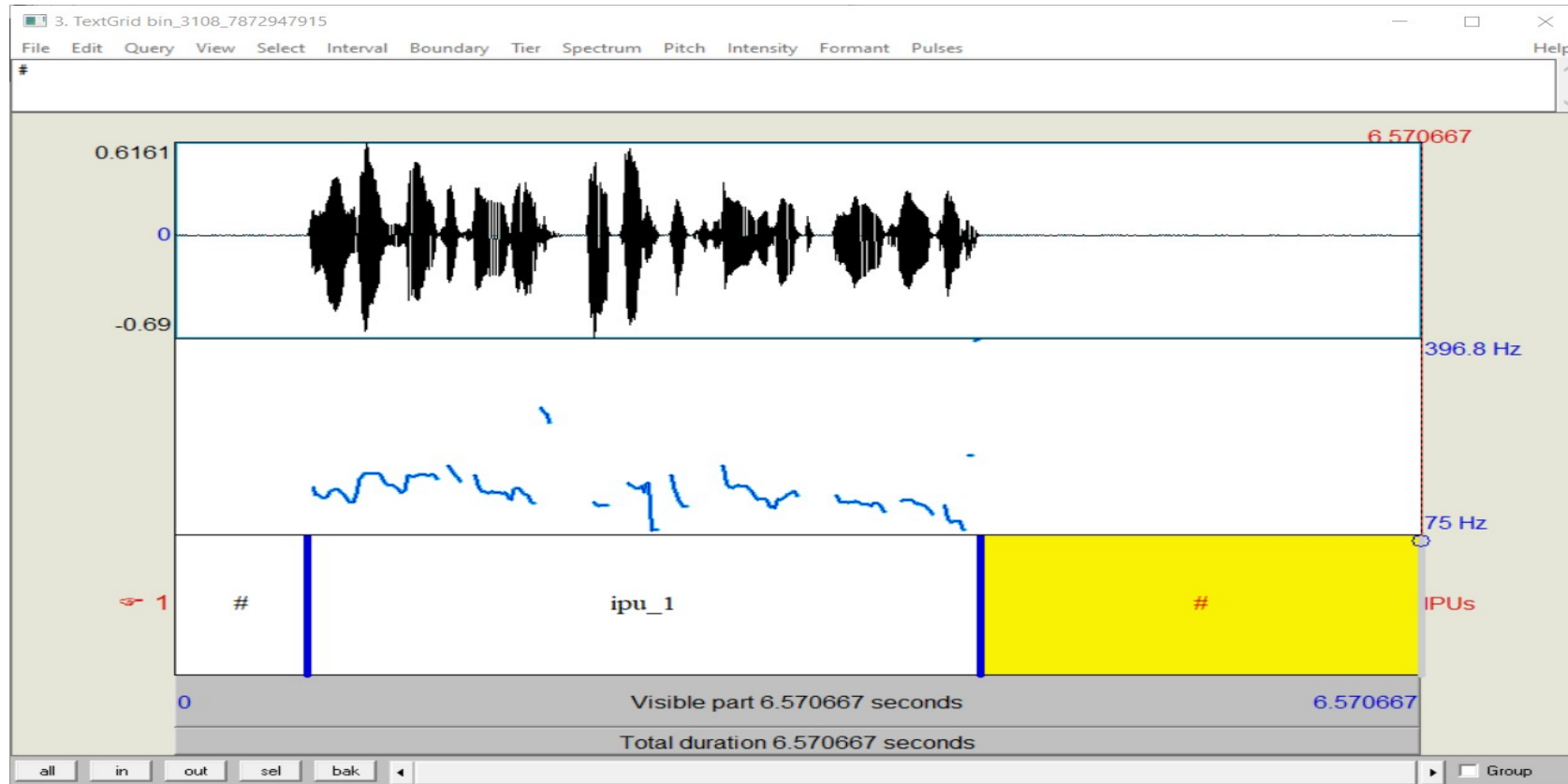
Bengali speech data analysis using SPPAS

- Search for IPU (Inter-Pausal Units) - A semi-automatic annotation process. Output is an annotated file with one tier named “IPUs”. Silence intervals are labelled with the “#” symbol, and IPUs intervals are labelled “ipu_” followed by the IPU number.



Figure 1: SPPAS window showing the IPU segmentation in speech waveform

- IPU segmentation in Praat textgrid



Perform manually the orthographic transcription

- Text Normalization - Unrestricted text with a variety of “non-standard” token types such as digit sequences, words, acronyms and letter sequences in all capitals, mixed case words, abbreviations, roman numerals, URL’s and e-mail addresses, needs to be normalized or rewritten using ordinary words.

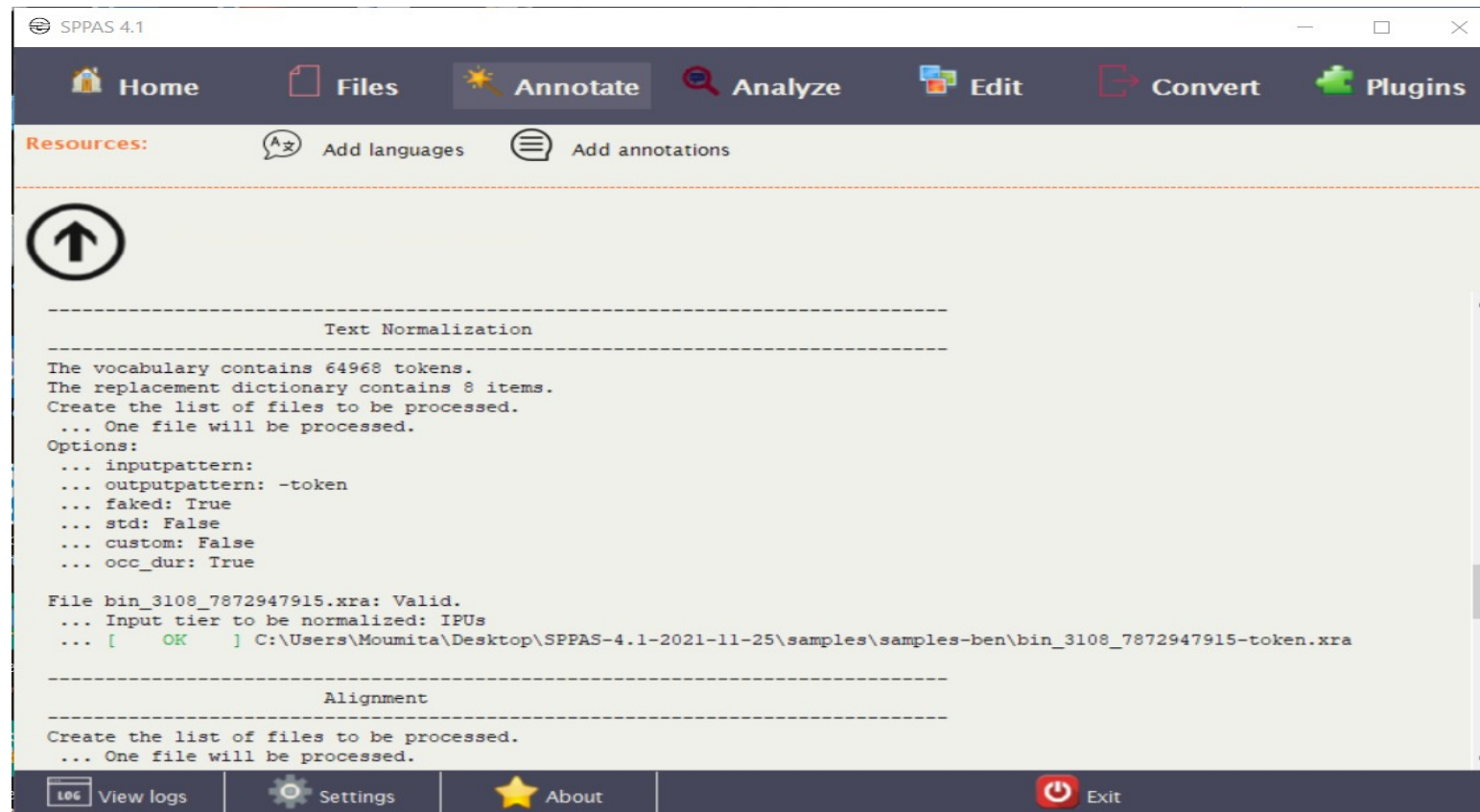


Figure 3: SPPAS annotation window showing the output results of text normalization

- **Phonetization** – Process of representing sounds with phonetic signs that produces a file with “-phon” appended to its name, i.e. “oriana1-phon.xra” for the previous example. This file contains only one tier with the resulting phonetization and with name “Phones”.
- **Alignment** – Alignment, also called phonetic segmentation, is the process of aligning speech with its corresponding transcription at the phone level.

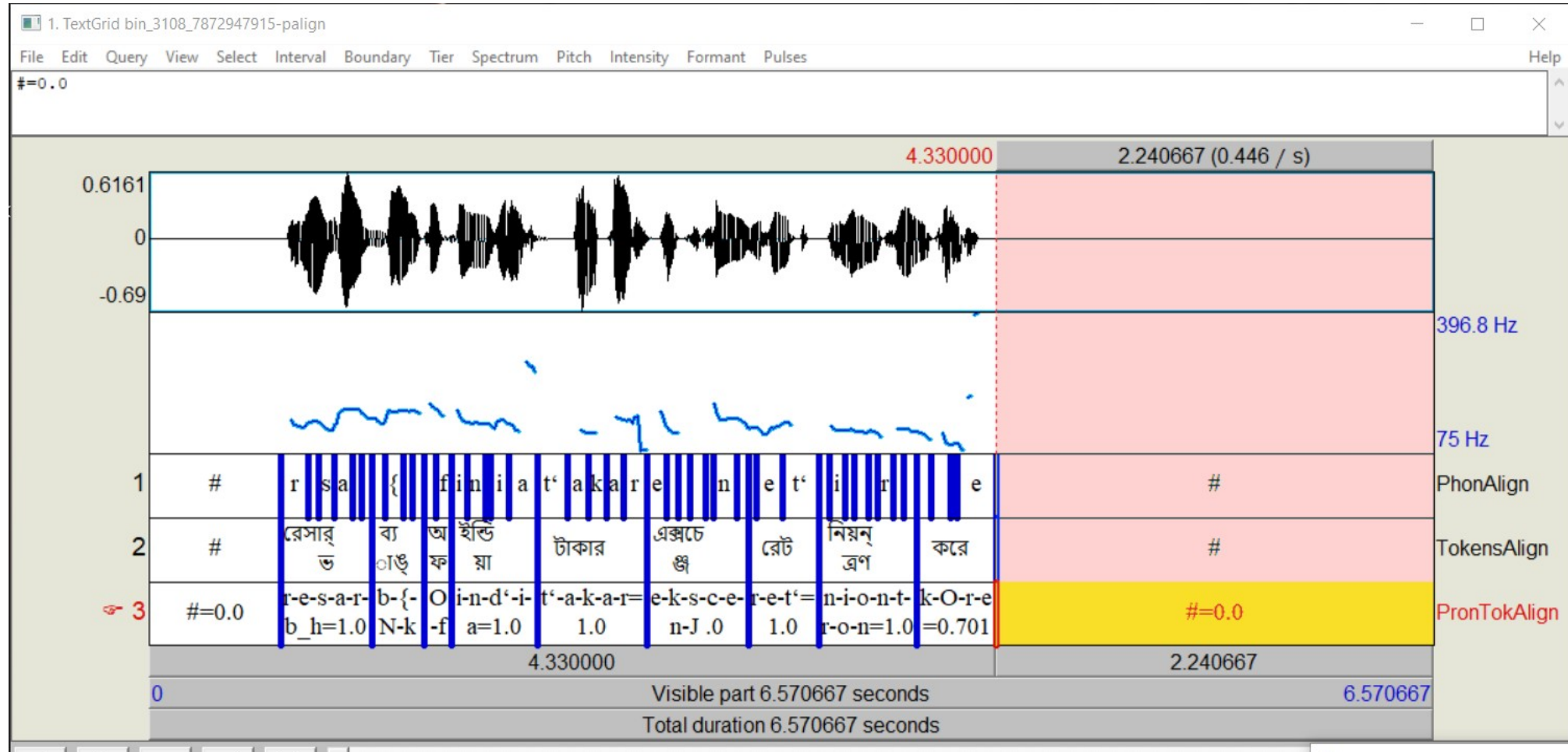


Figure 4: Praat window shows the final phonetized and aligned output results of a Bengali utterance.

- Descriptive statistical analysis – It allows to estimate the number of occurrences, the duration, etc. of the annotations of the analyzed tokens, and allows to save result in CSV (for Excel, OpenOffice, R, MatLab,...).

Stats View

View descriptive stats of tiers

N-gram: 1

Annotation label tags:

 Use only the tag with the best sco

 Use alternative tags

Annotation durations:

 Use only midpoint value

 Add the radius value

 Deduct the radius value

Summary Occurrences Total durations Mean durations Median durations Std dev. durations

Selected statistical distribution: summary

Label	Occurrences	Total durations	Mean durations	Median durations	Std dev. duration
0	138	0.0	0.0	0.0	0.0
0.5294117647058824	1	0.6779	0.6779	0.6779	0.0
0.7991071428571429	1	2.2407	2.2407	2.2407	0.0
1	117	2.9186	0.0249	0.0	0.215
1005	1	0.0	0.0	0.0	0.0
1005.0526315789474	1	0.19	0.19	0.19	0.0
1017	1	0.0	0.0	0.0	0.0
1026	1	0.0	0.0	0.0	0.0
1048	1	0.0	0.0	0.0	0.0
105	1	0.0	0.0	0.0	0.0
1056	1	0.0	0.0	0.0	0.0
1078	1	0.0	0.0	0.0	0.0
1093	1	0.0	0.0	0.0	0.0
11	4	0.0	0.0	0.0	0.0
1117	1	0.0	0.0	0.0	0.0
1133	1	0.2021	0.2021	0.2021	0.0
1142	1	0.0	0.0	0.0	0.0
1149	1	0.0	0.0	0.0	0.0
1150.5	1	0.08	0.08	0.08	0.0
1160	1	0.0	0.0	0.0	0.0
1172	1	0.0	0.0	0.0	0.0
1192	1	0.0	0.0	0.0	0.0
1193	1	0.0	0.0	0.0	0.0

Save Okay

Figure 5: SPPAS window showing the statistical analysis output results of durational measurements.

References

- Brigitte Bigi (2015). SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech. In “the Phonetician” - International Society of Phonetic Sciences, ISSN 0741-6164, Number 111-112 / 2015-I-II, pages 54-69.
- Brigitte Bigi, Daniel Hirst (2013). What’s new in SPPAS 1.5?, Tools and Resources for the Analysis of Speech Prosody, Aix-en-Provence, France, pp. 62-65.
- Brigitte Bigi (2012). SPPAS: a tool for the phonetic segmentations of Speech, The eight international conference on Language Resources and Evaluation, Istanbul (Turkey), pages 1748-1755, ISBN 978-2-9517408-7-7.
- Brigitte Bigi, Daniel Hirst (2012) SPEECH PHONETIZATION ALIGNMENT AND SYLLABIFICATION (SPPAS): A TOOL FOR THE AUTOMATIC ANALYSIS OF SPEECH PROSODY, Speech Prosody, Tongji University Press, ISBN 978-7-5608-4869-3, pages 19-22, Shanghai (China).
- Brigitte Bigi, Abiola S. Oyelere, Bernard Caron (2020). Resources for Automated Speech Segmentation of the African Language Naija (Nigerian Pidgin). Human Language Technologies Challenges for Computer Science and Linguistics, LNAI 12598, pp. 164-173, Springer, Heidelberg.
- Mélanie Lancien, Marie-Hélène Côté, Brigitte Bigi (2020). Developing Resources for Automated Speech Processing of Quebec French. In Proceedings of The 12th Language Resources and Evaluation Conference, pp. 5323–5328, Marseille, France.
- Keshan Sodimana, Knot Pipatsrisawat, Linne Ha, Martin Jansche, Oddur Kjartansson, Pasindu De Silva and Supheakmungkol Sarin (2018). A Step-by-Step Process for Building TTS Voices Using Open Source Data and Framework for Bangla, Javanese, Khmer, Nepali, Sinhala, and Sundanese. Proc. The 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU), Gurugram, India, pp. 66-70.
- <http://www.sppas.org/>
- P. Boersma, and D. Weenink, Praat: a system for doing phonetics by computer. (Version 6.2.01) Available at <http://www.praat.org>.

Thank you!