Introducing the SEA_AP: an Enhanced Tool for Automatic Prosodic Analysis

Marta Martínez*, Rocío Varela*, Carmen García-Mateo*, Elisa Fernández Rei, Adela Martínez Calvo*****

*AtlantTIC Research Center, Escola de Enxeñaría de Telecomunicación Universidade de Vigo, Campus As Lagoas Marcosende 36310 Vigo (Spain)

**Instituto da Lingua Galega, Universidade de Santiago de Compostela, Praza da Universidade, 4, 15782 Santiago de Compostela (Spain)

***Servizo de Consultoría Estadística, Dep. de Estatística e Investigación Operativa, Universidade de Santiago de Compostela, Facultade de Matemáticas, 15782 Santiago de Compostela (Spain)

E-mail: carmen.garcia@uvigo.es, mmartinez@gts.uvigo.es, rvarela@gts.uvigo.es, elisa.fernandez@usc.es, adela.martinez@usc.es

Abstract

SEA_AP (*Segmentador e Etiquetador Automático para Análise Prosódica, Automatic Segmentation and Labelling for Prosodic Analysis*) toolkit is an application that performs audio segmentation and labelling to create a TextGrid file which will be used to launch a prosodic analysis using Praat. In this paper, we want to describe the improved functionality of the tool achieved by adding a dialectometric analysis module using R scripts. The dialectometric analysis includes computing correlations among F0 curves and it obtains prosodic distances among the different variables of interest (location, speaker, structure, etc.). The dialectometric analysis requires large databases in order to be adequately computed, and automatic segmentation and labelling can create them thanks to a procedure less costly than the manual alternative. Thus, the integration of these tools into the SEA_AP allows to propose a distribution of geoprosodic areas by means of a quantitative method, which completes the traditional dialectological point of view.

The current version of the SEA_AP toolkit is capable of analysing Galician, Spanish and Brazilian Portuguese data, and hence the distances between several prosodic linguistic varieties can be measured at present.

Keywords: Speech segmentation, labelling, prosody, dialectometry

1. Introduction

Research activities produce an important amount of resources and tools which are usually unfocused and fragmented, since they are developed by different groups all over the world and implement partial solutions. It is also common that different research areas can help one another to reach better results based on the resources they share. As an example, in CORILGA project (García-Mateo et al., 2014), linguistics and speech processing had reached improvements for both research areas. Eventually, this situation turns into the need of unifying the material in compact code packages provided with user interfaces which will make the research progress also useful for the other specialization areas.

Taking into account this situation, the SEA_AP toolkit (*Segmentador e Etiquetador Automático para Análise Prosódica*) (Martínez et al., 2015) was developed. This software is an automatic segmentation and labelling tool which includes prosodic analysis scripts. In its first release (Martínez et al., 2015), it was prepared to get audio and transcription files as input and generate automatic timestamps which were used to create automatically a Praat TextGrid (Boersma & Weenik, 2013) to perform prosodic analysis. Besides, the tool seeks to integrate some tools used by linguists in dialectological oriented projects such as the AMPER project (Contini et al., 2002: 227-230; Martínez et al., 2003-2015).

One of the most important points of the prosodic analysis is the study of linguistic variations among different dialects. Results revealed by research on prosodic dialect variation are crucial for studying both diachronic prosody and those mechanisms taking part in prosodic change. One of the main issues defining geoprosodic research —and more generally, geolinguistics— is how to demarcate dialects, i.e. linguistic varieties sharing the same intonation that can therefore be considered geoprosodic varieties. Dialectometry offers a quantitative method enabling us to determine different geoprosodic areas. These areas are established according to a certain degree of shared similarities (see Fernández Rei and Martínez Calvo, 2014). This new approach distances itself from traditional dialectology, which used to establish a priori linguistic features that would then be used to determine the boundaries between different geographical varieties.

Therefore, the tool automates repetitive and arduous tasks, enabling the researcher to focus its efforts in the analysis and interpretation of data. Furthermore, this automation allows to compile large amount of data, which are essential for the statistical analysis carried out by the dialectometric method.

After a review of SEA_AP basic functionality, this paper introduces a new module which consists on a set of R scripts to perform dialectometric analysis.

2. SEA_AP Toolkit

SEA_AP tool is a desktop application for Windows that aligns audio and text without time marks and provides a segmentation with time labels corresponding to words, syllables or phones. Right after the segmentation phase, a prosodic study of the aligned data is launched using Praat scripts and interacting with the user for manual corrections or selection among the available analysis.

In order to start, the user must specify the audio folder, the text folder, the output folder where the TextGrid files will be generated and the directory of prosodic results. Once the user have clicked into the align button, segmentation and labelling tasks will start.
Afterword, prosodic analysis will be performed. The main goal is to extract the principal parameters of prosody. Its results are stored in text files.
Finally, a statistical study of the result obtained is carried out.
The complete system is illustrated as a block diagram in Figure 1.

2.1 Segmentation and labelling module

The first step to perform the alignment task consists in obtaining a phonetic transcription of the input text files using an automatic tool. This involves linguistic processing, morphological and syntactic analysis tasks. After that, the correspondent phone sequence is assigned to each grapheme. This task has been performed with the software Cotovía (Rodríguez Banga et al., 2012), for Galician and Spanish languages, and with the grapheme to phone converter developed under the project FalaBrasil (Nelson et al., 2010) for Brazilian Portuguese.
Next phase consists in the extraction of acoustic features from the audio file. The type of features that has been used in this step is Mel-Frequency Cepstral Coefficients, MFCCs.
Acoustic models need to be generated to perform a recognition phase. This had been modelled based on Hidden Markov Models (HMMs) using the HTK Toolkit (Young et al., 1995). This tool uses inductive learning based on labelled databases during a training phase. For this purpose, acoustic models have been trained with this databases according to the language:
- Spanish: trained with TCSTAR (Docío-Fernandez et al., 2006) which contains European and Spanish Parliament recordings with their respective transcriptions.
- Galician: trained with Trascrigal database (García-Mateo et al. 2004) which contains 31 hours of news recordings from regional television with transcriptions.
The acoustic model used for Brazilian Portuguese have been obtained from the repository of the project FalaBrasil, where it is available as a public resource.
In addition, a linguistic modelling block is needed to create a network which tries to predict the most probable word. In this particular case, we will perform forced alignment, which means that the data used to create this network is the sentence from the input text file since that we certainly know the most probable word to be predicted.
Eventually, this data is passed to the recognition block that combines all the information and assigns a temporal interval to each considered segment as well as a certain probability. This final recognition phase is performed using HTK Toolkit.
In this step, segmentation and vowel labelling of speech is obtained, which will be the basis on which the prosodic analysis is sustained.

2.2 Prosodic analysis module

The study of prosody focuses on the description of the evolution of the fundamental frequency, the duration and the intensity linked to the sound chains which convey linguistic meaning.
Prosody study is accomplished by using Praat scripts implemented by the Laboratory of Phonetics from the University of Barcelona (Elvira-García *et al*., 2014), under the AMPER project (Rilliard, 2013) and by the Multimedia Technologies Group (GTM) of the University of Vigo.
By using these scripts, users can analyse the waveform, the spectrogram, the fundamental frequency curve, the values of three points in each TextGrid interval, the duration, the intensity and it can also extract labels for each interval and the data position in relation with the accent.
Moreover, some scripts let the user make modifications into TextGrid files editing or removing tiers, creating empty TextGrids to every file in an audio folder or the generation of words and syllables tiers from the phonetic tier, and the generation of a new tier with vowel information, since they are very important to prosodic studies. In addition, the script *create_pictures.praat* (Elvira-García, et al., 2014) also allows the user to generate graphics that contain different prosodic information. An example is shown in Figure 2.

3. Dialectometric module

3.1 Description of the dialectometric scripts

The dialectometric module was developed with the statistical software R, and it was incorporated into the SEA_AP toolkit in order to make a dialectometric analysis of the generated prosodic files with the aim of performing an analysis of prosodic variation which is one of the fields of study of the prosody. The dialectometric methodology will allow us to establish geoprosodic areas, that is, geographic areas with similar intonation patterns. Furthermore, it will also enable the establishment of the degree of proximity/distance between the different intonation patterns used by the speakers, so that we will be able to study the intonation variability of a speaker or of a given place or region, for instance. This methodology also provides us with the degree of proximity/distance between the intonation patterns gathered in the different survey points in order to verify which varieties are prosodically closer and which are more distant.
The input data of the dialectometric module are the fundamental frequency values generated by the previously introduced modules, and stored as text files (see an example of this kind of files in Figure 3). After reading the F0 data from the text files, the functionalities included in this module allow the researcher to compute correlations among F0 curves and to obtain prosodic distances among the different locations (speakers or another variable of interest) existing in the prosodic data. An example of this kind of outputs of the module is shown in Figure 4. Next, multivariate statistical techniques such as multidimensional scaling (MDS) and cluster analysis can be applied to the table of prosodic distance. Specifically,

using the cluster analysis, the researcher could detect the existence of clusters of locations (speakers or another variable of interest) according to their closeness in terms of the measure of the prosodic distance. Furthermore, all the statistical techniques of the dialectometric module generate both numerical and graphical outputs that summarize the main results of the analysis. One of the graphical outputs provided allows mapping the different geolinguistic areas produced by the analysis of the prosodic data obtained. An example of this mapping is shown in Figure 5.

3.2 Dialectometric scripts integration

Once the prosodic analysis module is finalised, the dialectometric module will start by launching the R console where the user can obtain the results from the performance of the script. Results can also be stored into text files for further analysis and they will be located in the output folder.

4. Cases of study

In this section we describe possible scenarios of use.

Many researches are still being done on the study of languages. In recent years we have become increasingly aware of the importance of teaching prosody when learning a foreign language. The tool could be used to analyse how effective a specific teaching method of pronunciation is and to compare the effectiveness of different teaching methods of prosody in foreign languages acquisition. The importance of studying this feature lies on intonation as it is crucial for communication, changing the intonation can completely change the meaning or even without intonation, it is impossible to understand the expressions. Therefore, this means that teaching prosody is very significant when acquiring a language different from the mother tongue. The tool can also be used to study how it affects the prosodic distance between the mother tongue and the target language and the purpose will be to measure how this feature affects the acquisition of new languages for the different methods of teaching prosody and it will allow to identify which method is appropriate in each case in terms of the prosodic distance.

Furthermore, the application could be used to perform more accurate prosodic division areas by including new data which will be processed more quickly by the tool. Thus, it allows to group prosodic features into smaller areas and consequently it represents each one of the areas with greater accuracy.

5. Conclusions and further work

SEA_AP goal is to automatize the laborious manual work of segmentation and labelling tasks. This automatization allows the users to effortlessly generate the large prosodic datasets which are required for a reliable dialectometric analysis. Once that a prosodic dataset is created, the dialectometric module, incorporated into the SEA_AP toolkit, can analyse the data using statistical techniques and compute measures of the prosodic distance among locations, speakers or another variable of interest which is presented in the data. Furthermore, cluster analysis can be applied by the researchers in order to detect geoprosodic areas based on the obtained prosodic distances.

A possible line of research to continue with this project would be to make an evaluation of the accuracy in the alignment results obtained with SEA_AP making a comparison with a manual labelled database.

Due to the fact that alignment accuracy depends on the quality of the forced alignment and acoustic models, it seems possible to improve the models using aligned data to get better alignment results.

Finally, another possible line of work is to expand the tool to other languages.

6. Acknowledgements

This work has been supported by the Galician Regional Government (CN2011/019, CN2012/160, CN2012/179, GRC2013/40), the European Regional Development Fund (ERDF) and the Spanish Government ('SpeechTech4All Project' TEC2012-38939-C03- 01).

7. Bibliographical References

Boersma, Paul & Weenik, David (2013): Praat: doing phonetics by computer [Computer program]. Version 5.4.05, downloaded: 16 of February of 2015 de http://www.praat.org/

Contini M., Lai J.-P., Romano A., Roullet S., de Castro Moutinho L., Coimbra R. L., Pereira Bendiha U. et Ruivo S. S. (2002). "Un Projet d'Atlas Multimédia Prosodique de l'Espace Roman ", dans B. Bel et I. Marlien (éds), Proceedings of the Ist International Conference on Speech Prosody (Aix-en-Provence, 11-13 avril 2002), p. 227-230.

Docio-Fernandez L., Cardenal-Lopez A., and García-Mateo, C. (2006). "TC-STAR 2006 Automatic Speech Recognition Evaluation: The UVIGO System". In: TC-STAR Workshop on Speech-to-Speech Translation.

Elvira García, W., Roseano, P., & Fernández Planas, A. M. (2014). Praat scripts. Downloaded on June 1, 2015 in http://stel.ub.edu/labfon/en

Fernández Rei, Elisa & Martínez Calvo, Adela (2014). "Proposta de análise dialectométrica da prosodia galega". In Textos Selecionados XXIX Encontro Nacional da Associação Portuguesa de Linguística, Porto, Portugal, 2013, pages 249-254.

Garcia-Mateo C., J. Dieguez-Tirado, A. Cardenal-Lopez, and L. Docio-Fernandez. (2004). "Transcrigal: A bilingual system for automatic indexing of broadcast news". In Proc. Int. Conf. on Language Resources and Evaluation, volume 6, Lisbon, Portugal, May, pages 2061–2064.

García-Mateo C. & A. Cardenal & X. L. Regueira Fernández & E. Fernández Rei & M. Martínez & R. Seara & R. Varela & N. Basanta Llanes (2014). "CORILGA: a Galician Multilevel Annotated Speech Corpus for Linguistic Analysis". 9th Language Resources and Evaluation Conference (LREC 2014). Reykjavik, 26-31 May 2014.

Martínez Celdrán, Eugenio & Fernández Planas, Ana Ma. (coords). 2003-2015. Atlas Multimèdia de la

Prosòdia de l'Espai Romànic. http://stel.ub.edu/labfon/amper/cast/index_ampercat.html

Martinez, M. & Varela, R. & López, P. & Docío, L.(2015):"SEA_AP: Segmentador e Etiquetador Automático para a Análise Prosódica", Colóquio Internacional de Geoprosódia do Português e do Galego, Aveiro 2015.

Nelson Neto, Patrick Silva, Aldebaro Klautau, Isabel Trancoso (2010). "Free tools and resources for Brazilian Portuguese speech recognition". Journal of the Brazilian Computer Society (Online), 2010, http://link.springer.com/article/10.1007%2Fs13173-010-0023-1

Rodríguez Banga, E, García Mateo, C, Méndez Pazó, FJ, González, M, Magariños Iglesias, C (2012). "Cotovía: an open source TTS for Galician and Spanish". In IberSPEECH 2012 "VII Jornadas en Tecnología del Habla and III Iberian SLTech Workshop".

Rilliard Albert (2013): "Metodoloxía Cuantitativa para a Medida das Distancias Prosódicas" "Xornadas de Dialectoloxía Perceptiva". Available in: http://ilg.usc.es/tecandali/Descargas/AlbertRilliard.pdf

Young, Steve & Gunnar Evermann & Mark Gales & Thomas Hain & Dan Kershaw & Gareth Moore & Julian Odell & Dave Ollason & Dan Povey &Valtcho Valtchev & Phil Woodland (1995) HTK Book. Available in: http://htk.eng.cam.ac.uk/docs/docs.shtml

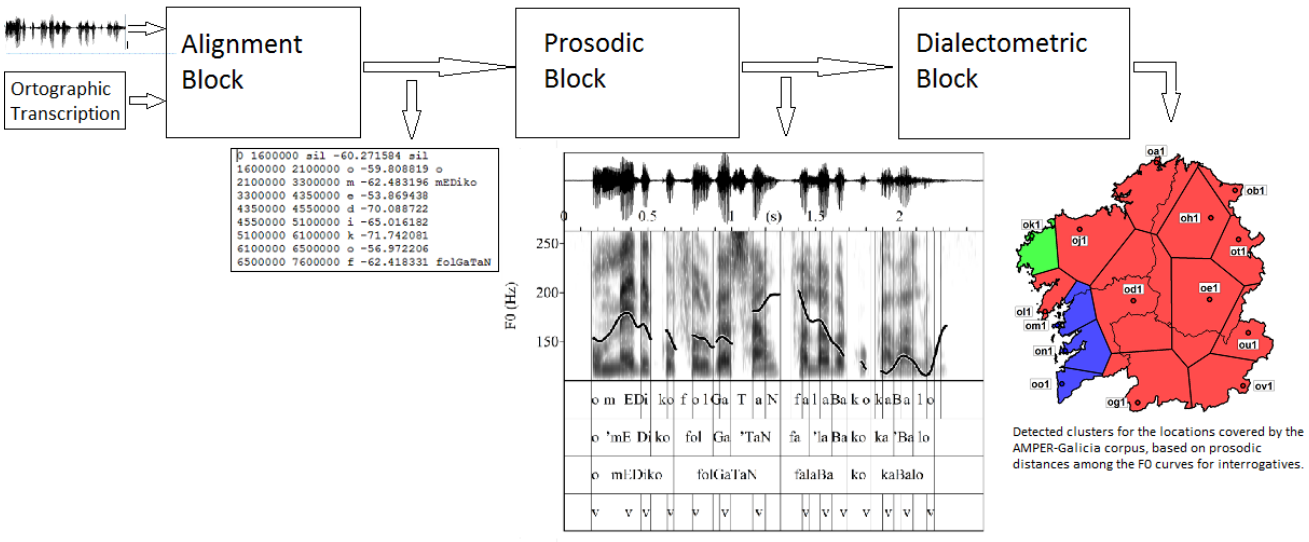

Figure1: Block Diagram

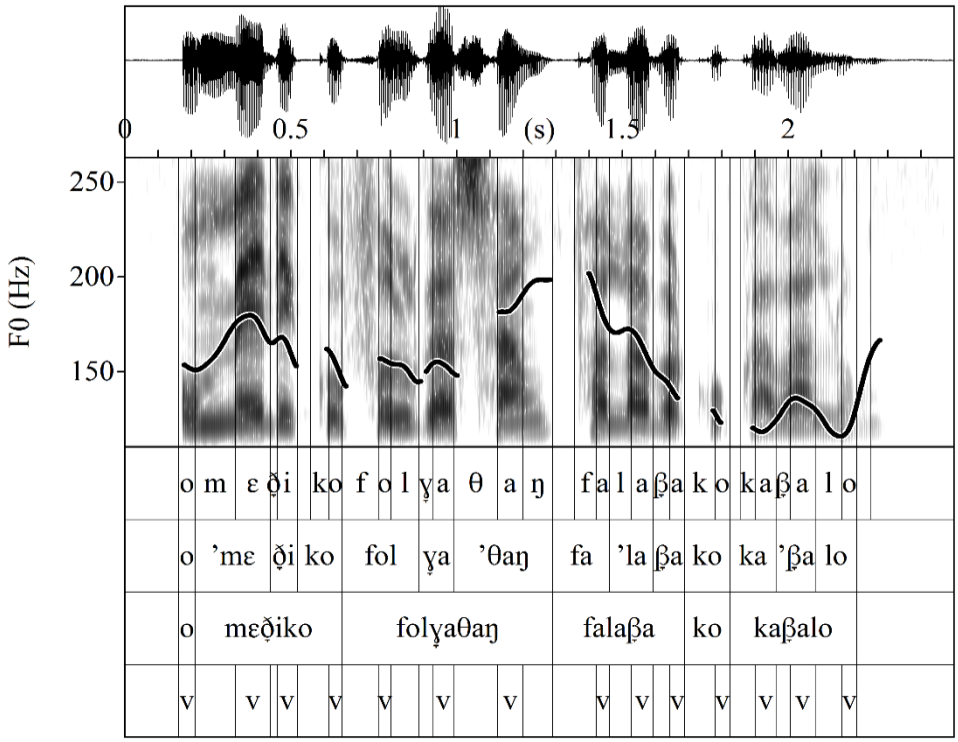

Figure2: Example of picture generated by the SEA_AP toolkit, which contains a waveform, a spectrogram, a F0 track and the content of the tiers of the TextGrid file.

	duration [ms]	energy [dB]	fo1	fo2	fo3
[Hz]					
1	50	81	152	152	150
2	60	80	149	174	158
3	40	79	169	156	151
4	35	80	157	157	153
5	65	83	159	154	155
6	75	80	182	176	189
7	40	81	198	183	172
8	65	81	179	168	153
9	45	80	148	141	129
10	45	72	139	124	124
11	65	74	123	122	122
12	75	77	133	137	134
13	45	67	129	130	132

values at:
2560 2960 3360 7280 7760 8240 9760 10080 10400 12160 12440 12720 14720
15240 15760 17840 18440 19040 22560 22880 23200 24240 24760 25280 26080
26440 26800 28240 28600 28960 30160 30680 31200 31840 32440 33040 34320
34680 35040

Figure3: Example of text file generated by the SEA_AP toolkit, which contains the fundamental frequency data required by the dialectometric module.

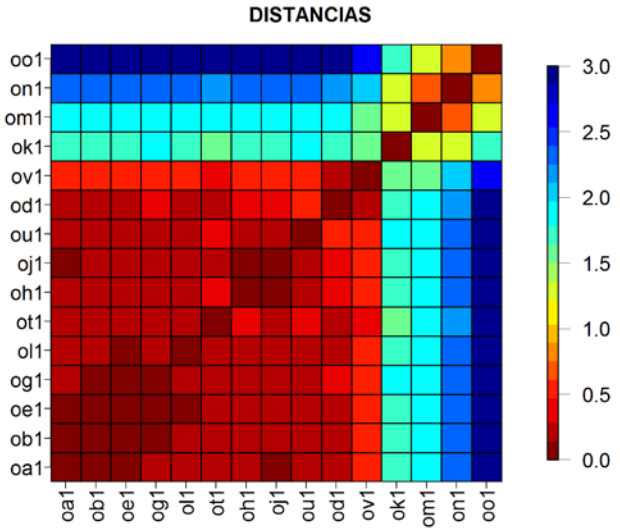

Figure4: Example of dialectometric module output: prosodic distances among locations covered by the AMPER-Galicia corpus, based on the analysis of the F0 curves for interrogatives.

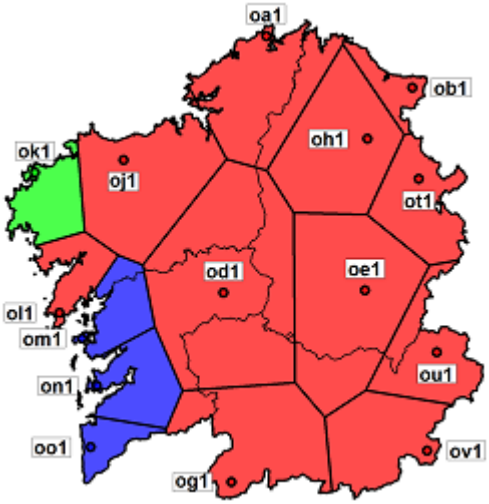

Figure5: Result of dialectometric module. Detected clusters for the locations covered by the AMPER-Galicia corpus, based on analysis of the F0 curves for interrogatives. The colour of the polygon associated with each location indicates to which of the three detected clusters it belongs: cluster 1 (red), cluster 2 (blue) or cluster 3 (green).