DETERMINING SPEAKERS' ORIGIN ON THE BASE OF THEIR INTONATION: A PRELIMINARY STUDY

Wendy Elvira-García[1], Paolo Roseano[2,3], Ana Ma. Fernández Planas[2]

[1]Universidad Nacional de Educación a Distancia, [2]Universitat de Barcelona, [3]University of South Africa
welvira@flog.uned.es, paolo.roseano@ub.edu, anamariafernandez@ub.edu

ABSTRACT

Este trabajo presenta AMPER_Forensic, una herramienta diseñada para ayudar a los lingüistas en la tarea de identificar el origen de un hablante. La herramienta usa datos exclusivamente prosódicos y propone al usuario una lista de lugares en los que se usan contornos entonativos parecidos a los que aparecen en la grabación objeto de estudio. AMPER_Forensic usa técnicas conocidas que se utilizan normalmente en estudios de dialectometría, aplicándolas a la tarea de atribución de dialecto, una parte del LADO que ha ganado popularidad en los últimos años.

Palabras clave: entonación, fonética forense, fonética judicial, AMPER

This paper presents AMPER_Forensic, a tool aimed at helping linguists in the task of assessing speaker's origin. The tool uses prosodic data exclusively and suggests to the user a list of places that use intonational patterns similar to those who appear in the target recording. AMPER_Forensic uses well-known dialectometric techniques and applies them to the dialect attribution task, which is a part of LADO that has been gaining popularity in recent years.

Keywords: intonation, forensic phonetics, AMPER

1. INTRODUCTION

Due to its importance in court cases, language analysis for the determination of origin (LADO), and specifically dialect attribution, has been the object of increasing research (Allen, DeLima, Freed, & Nielsen, 2019). However, determining speakers' origin is a difficult task even for trained linguists (Cambier-Langeveld, 2010).

For this reason, recent studies use *automatic* methods in order to identify dialects (Brown & Watt, 2014; Ferragne & Pellegrino, 2007; Hanani, Russell, & Carey, 2013; Vidya Prasad et al., 2019). No matter if the above-mentioned studies use textual data or spoken data, they tend to use a holistic approach, without giving the expert a clear indication concerning which parameters contribute the most to the result of the analysis they carry out. Only in few cases, automatic systems act "human-like" insofar as they choose some parameters that are believed to be qualitatively more significant than others. For example, Biadsy's (2011) system performs a classification relying only on phones that are known to be realized differently in different dialects.

On the other hand, methods based on *human* expertise do not perform holistic analyses but choose some parameters that, according to the knowledge of the experts, can perform better than others can. Van Bezooijen and Gooskens (1999) explore the contribution of different linguistic levels to the identification of language varieties by human listeners and they conclude that "pronunciation" (sounds and prosody altogether) is more significant than the rest of levels. However, despite the existence of studies on the contribution of intonation to speaker identification (Cicres, 2007), the research about the contribution of intonation to dialect detection is not abundant.

This paper aims at testing whether prosody can be an important feature in semi-automatic origin determination. In order to do so, we have designed a tool that can help experts in their task of determining speaker's origin in Romance languages.

2. MATERIALS

2.1. Corpora

Determining speaker's origin requires a dataset that includes data from several survey point. To date, there are two intonational databases that could serve that purpose for Romance languages: the *Interactive Atlas of Romance Intonation* (IARI; Prieto, Borràs-Comes, & Roseano, 2010-2014) and the *Atlas Multimédia de la Prosodie de l'Espace Roman* (AMPER; Contini, 1992; Romano & Contini, 2001; Romano, 2003). This paper uses the latest.

The AMPER database consists of several corpus containing speech samples of different formality grades, ranging from spontaneous speech to laboratory read speech, including map tasks and discourse completion tasks.

In order to work with intonation, this study requires that the F0 contours are aligned among all samples. Therefore, the read speech corpus has been chosen. Specifically, this study uses read yes-no questions in three languages: Italian, Friulian (Roseano & Fernández Planas, 2009-2013) and Catalan (Martínez Celdrán & Fernández Planas, 2003-2019).

Every question of our corpus is formed by three constituents with lexical stress in the same positions (all of them are paroxytones). Examples are provided in (1) for Italian, (2) for Friulan and (3) for Catalan.

(1) It. *La bambina mangiava la banana?*
(2) Fri. *La ghitare si sunie cun dolcece?*
(3) Cat. *El copista no porta la caputxa?*

2.2 Speakers

2.2.1. Accents database

The accents database used in this study includes recordings from 18 speakers of catalan Catalan (i.e. 2 from each of the following survey points: Alicante, Andorra, Barcelona, Castelló de la Plana, Fraga, Girona, Tarragona, Tortosa, and Valencia), 8 speakers of Friulian (2 in each of the following survey points: Agrons, Beivars, Gradisca and Tesis), and 5 speakers of Italian from the following points: Venezia, Arezzo, Siena, Firenze, and Perugia.

In total 31 speakers of 18 survey points were included in the study.

2.2.2. Test database

In order to test if the attribution of accent performs correctly, we need a set of recordings that can be compared witch the accent database. Therefore, we chose a set of recordings that could serve as "accent unknown recording" (AUR), which we will call test database.

In order to create the test database, we have used the following approach. In the survey points where we recorded two speakers, each recording has been excluded once from the training set and has been analysed as if it were the AUR. This allows us to test the performance of the program 26 times with different recordings.

2.3 Extraction and processing of prosodic data

The tool that we are presenting uses prosodic data, therefore sound files must be pre-processed in order to be correctly analysed.

To that end, the recordings have been analysed with the program AMPER06 that allows the user to extract and stylize F0 contours (López Bobo et al., 2007). The program saves in a txt, for each sentence, three F0 values in semitones per syllable (measures taken at the starting point, at the mid point, and at the end of each vowel), a value of vowel duration per syllable and the value of the mean intensity of each syllable.

3. METHODS

The tool created for helping the linguist in the task of determining speakers origin is a Python script called AMPER_Forensic. Its main goal is providing a rank of survey points that match the intonational contour used in the AUR. In order to do so, AMPER_Forensic takes the intonational contour of a given recording and compares it to the contours of previously recorded sentences that show prototypical accents of each region. In following sections, the pipeline of AMPER_Forensic is described in detail.

3.1 Input

The input of the program, as we have stated in section 2, is a set of txt files containing prosodic data. Such txt files (see Figure 1 for an example) are generated with AMPER06, but the program can be adapted to any kind of input as long as the input keeps the syllable/stress position correspondence in all the languages/varieties used in the dataset.

```
C:\Amper\ficherostxtSt\8001005p0p0p0a0u0e03x6xAx11i1.txt
     size: 29609
29-Oct-2010

zona        duration [ms]          energy [dB]            fo1
     fo2   fo3 [ST]

 1           31         101         -0.32 -0.07 0.25
 2           47         104          1.71  2.58  3.33
 3           96         104          3.51  0.73 -0.57
 4           47          99         -1.52 -2.02 -2.47
 5           47          93         -2.43 -1.61 -1.08
 6           51          99          1.42  2.51  3.46
 7           68          99          3.04  2.65  2.15
 8           46          96          0.49 -0.07 -0.45
 9           44          99         -2.70 -2.70 -2.62
10           96         100         -2.91 -2.91 -2.91
11          114          98         -1.90  0.33  3.13

values at:
1687 1935 2182 3095 3473 3851 5389 6158 6927 7527 7905 8283
10030 10408 10786 12480 12884 13288 14149 14696 15244 16573
16938 17303 18503 18855 19206 21735 22504 23274 25516 26428
27341
```

Figure 1. Example of an input file. It includes data of F0 in semitones, duration and intensity of the sentence.

3.2 Computing intonation differences

Once the AUR sentence is loaded, the system compares it with the sentences in the database (in the case of this paper, the corpora explained in section 2.2.2). This comparison is performed by computing the similarities and differences between the AUR and the rest of them. In order to do so, the program computes Pearson correlation, as this is the metric that has proved to be more useful for measuring similarities among intonation contours (Hermes, 1998). The approach is similar to what is done in intonational dialectometry (Elvira-García et al., 2018).

Correlation values range from -1 to 1, where 1 is the perfect correlation. Thus, the higher values of correlation, the more likely that the two sentences analysed are from the same dialectal area.

4. RESULTS

The results of the program contain all the correlations of the AUR with the contours in the Accents database. Such results are offered to the user in two ways: numerically (in a txt) and visually (as depicted in Figure 2) ordered from highest correlation coefficient to lowest correlation coefficient. Therefore, survey points that share intonational patterns with the recording appear in the highest part of the rank (left side of Figure 2). On the contrary, cities with other patterns appear with lower correlation indices (in right part of the figure).

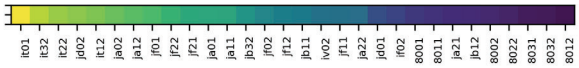

Figure 2. Heatmap of ordered correlations created by AMPER_Forensic.

For the dataset analysed, the results have correlated each AUR with recordings of the same linguistic area (Friulian, Catalan and Italian) in 100% of the cases. However, the speaker with the highest correlation with the AUR is not always a speaker from the same survey point. We have assessed classifications as correct when the two speakers of the same city have a correlation higher than 0.9 (this is true in 42% of the cases).

5. LIMITATIONS AND FUTURE WORK

The results obtained are satisfactory but not perfect, insofar as AMPER_Forensic identifies correctly to which large geoprosodic area the AUR belongs to, but most of the times it does not identify correctly the survey point. In this section, we try to explain possible reasons for that.

The poor performance of the tool for locality identification was not unexpected. In fact, it would be rather obvious for any dialectologist. Whereas some dialectal areas have different intonational patters from town to town, others, especially varieties with a strong standard or areas with great mobility are levelling their accents making them almost disappear. For those places, we could never achieve a good performance. If speakers do not have a “survey point accent”, neither a machine nor a linguist can find it (Dyer, 2002). This is most likely the case for many of the Catalan dialectal areas, since previous research has shown that cities like Andorra and Barcelona share the same intonational pattern in SVO yes-no questions (Fernández Planas et al., 2015).

However, this is not the only factor that could have had an effect on the outcome of the analysis we present in this paper. The approach adopted to test the data (see Section 2.2.1) has made that for each trial we had only a speaker of the same survey point as the AUR in the database. Therefore, if the two speakers of the same survey point used different intonational patters (which is something that can happen due to the fact that some variation is always observed in intonation), AMPER_Forensic has not been able to identify correctly the survey point. In order to overcome this shortcoming, more speakers

of each survey point should be included in the database.

Furthermore, a more comprehensive model should be created. In this paper, we have compared the contours one to one. Nevertheless, in an ideal scenario, when a survey point presents more than one intonational pattern, the AUR should be compared to every contour attested in the area. In order to do so, the most straightforward way is changing the approach and constructing a model that includes all the features attested in the dialect. For doing so, the next step we are working on is using a neural network classifier.

6. CONCLUSIONS

In this paper we have presented AMPER_Forensic, a tool aimed at helping linguists in the task of assessing speaker's origin using prosody. The results show that intonation can be used in dialect identification tasks insofar as AMPER_Forensic is able to assign an AUR to its geoprosodic area. However, the system does not to recognize the specific survey point from which an AUR comes from. This limitation can be explained by the natural intra-speaker and inter-speaker variation and can be solved by means of a more comprehensive classification method, like new machine learning techniques.

7. REFERENCES

Allen, C. O., DeLima, R., Freed, A. R., & Nielsen, R. L. (2019). *U.S. Patent Application No. 10/275,446*.

Biadsy, F. (2011). *Automatic dialect and accent recognition and its application to speech recognition*. PhD thesis, Columbia University.

Brown, G. and Watt D. 2014. Performance of a novel automatic accent classifier system using geographically proximate accents. Poster session *BAAP*, University of Oxford, 7th-9th April.

Cambier-Langeveld, T. (2010). The role of linguists and native speakers in language analysis for the determination of speaker origin. *International Journal of Speech, Language & the Law*, 17(1), 67-93.

Cicres, J. (2007). A*plicació de l'anàlisi de l'entonació i de l'alineació tonal a la identificació de parlants en fonètica forense*. PhD thesis, Universitat Pompeu Fabra.

Contini, M. (1992). Vers une géoprosodie romane. In G. Aurrekoetxea & X. Bidegain (Eds.). *Actas del Nazioarteko Dialektologia Biltzarra Agiriak,* Bilbao 1991. Bilbao: Publicaciones de la Real Academia de la Lengua Vasca, 83-109.

Dyer, J. (2002). 'We all speak the same round here': Dialect levelling in a Scottish-English Community. *Journal of Sociolinguistics*, 6(1), 99-116.

Elvira-García, W., Balocco, S., Roseano, P., & Fernández Planas, A. M. (2018). ProDis: A dialectometric tool for acoustic prosodic data. *Speech Communication* 97, 9-18.

Fernández Planas, A. M., Roseano, P., Elvira-García, W., Cerdà-Massó, R., Romera-Barrios, L., Carrera-Sabaté, J., Szmidt, D., Labraña, S., & Martínez Celdrán, E. (2015). Cap a un nou mapa dialectal del català? Consideracions a partir de dades prosòdiques tractades dialectomètricament. *Estudios de Fonética Experimental* 24, 257-286.

Ferragne, E., & Pellegrino, F. (2007). Automatic dialect identification. A study of British English. In Müller, C. (Ed.). *Speaker classification*. Berlin: Springer, 243-257.

Hanani, A., Russell, M., & Carey, M. (2013). Human and computer recognition of regional accents and ethnic groups for British English speech. *Computer Speech and Language* 27, 59-74.

Hermes, D. J. (1998). Measuring the perceptual similarity of pitch contours*. Journal of Speech, Language, and Hearing Research*, 41(1), 73-82.

López Bobo, M. J, Muñiz Cachón, C., Díaz Gómez, L. , Corral Blanco, N. Brezmes Alonso, D., & Alvarellos Pedrero, M. 2007. Análisis y representación de la entonación: Replanteamiento metodológico en el marco del proyecto AMPER. En Dorta, J. (Ed.). *La prosodia en el ámbito lingüístico románico.* Santa Cruz de Tenerife: La Página, 17-34.

Martínez Celdrán, E., & Fernández Planas, A. M. (Eds.) (2003-2019). *Atlas multimèdia de la prosòdia de l'Espai romànic*. http:/stel.ub.edu/labfon/amper/cast/index_ampercat.html (2019-12-12).

Prieto, P., Borràs-Comes, J., & Roseano, P. (Eds.) (2010-2014). *Interactive atlas of Romance intonation*. http://prosodia.upf.edu/iari/ (2019-12-12).

Romano, A. (2003). Un projet d'atlas multimédia prosodique de l'espace roman (AMPER). In Sánchez Miret, F. (Ed.). *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica.* Salamanca, 24-30 septiembre 2001, vol. I: Discursos inaugurales. Conferencias plenarias. Sección 1: Fonética y fonología. Sección 2: Morfología. Índices: Índice de autores, Índice general. Tübingen, Niemeyer, 279-294.

Romano, A., & Contini, M. (2001). Un progetto di atlante geoprosodico multimediale delle varietà linguistiche romanze. In Magno Caldognetto, E. & Cosi, P. (Eds.). *Multimodalità e multimedialità nella comunicazione. Atti delle XI Giornate di Studio del Gruppo di Fonetica Sperimentale dell'Associazione Italiana di Acustica*. Padova: Unipress, 121-126.

Roseano, P., & Fernández Planas, A. M. (Eds.) (2009-2013). *Atlant multimediâl de prosodie des varietâts romanichis*. http://stel.ub.edu/labfon/amper/friul/index.html/ (2019-04-04).

Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different

linguistic levels. *Journal of language and social psychology*, 18(1), 31-48.

Vidya Prasad, K., Akarsh, S., Vinayakumar, R., & Soman, K. P. (2019). A Deep Learning Approach for Similar Languages, Varieties and Dialects. *arXiv preprint* arXiv:1901.00297.

Acknowledgements. This paper has been developed within the framework of the project *Technologies derived from AMPER-CAT and analysis of complementary corpora* (FFI2015-64859-P) funded by the Spanish *Ministerio de Economía y Competitividad*. We thank the Faculty of Philology and Communication of the University of Barcelona for its financial support for the acquisition of the GPU (NVIDIA GTX 1060 6GB) that has been used in this study.